



PHD

The general practice research database as an alternative to registries for studying drug safety in pregnancy Anticonvulsants as a case study

Charlton, Rachel

Award date:
2012

Awarding institution:
University of Bath

[Link to publication](#)

Alternative formats

If you require this document in an alternative format, please contact:
openaccess@bath.ac.uk

Copyright of this thesis rests with the author. Access is subject to the above licence, if given. If no licence is specified above, original content in this thesis is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC-ND 4.0) Licence (<https://creativecommons.org/licenses/by-nc-nd/4.0/>). Any third-party copyright material present remains the property of its respective owner(s) and is licensed under its existing terms.

Take down policy

If you consider content within Bath's Research Portal to be in breach of UK law, please contact: openaccess@bath.ac.uk with the details. Your claim will be investigated and, where appropriate, the item will be removed from public view as soon as possible.



**The General Practice Research Database as an alternative
to registries for studying drug safety in pregnancy:
anticonvulsants as a case study**

Rachel Charlton

A thesis submitted for the degree of Doctor of Philosophy

University of Bath

Department of Pharmacy and Pharmacology

March 2012

COPYRIGHT

Attention is drawn to the fact that copyright of this thesis rests with the author. A copy of this thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright rests with the author and that they must not copy it or use material from it except as permitted by law or with the consent of the author.

This thesis may be made available for consultation within the University Library and may be photocopied or lent to other libraries for the purposes of consultation.

Signed

Date

Abstract

Background: In recent years, there has been an increase in the use of automated healthcare databases for drug safety in pregnancy evaluation; their suitability for this purpose needs to be evaluated.

Aim: To evaluate the utility of the United Kingdom's General Practice Research Database (GPRD) to act as an alternative to pregnancy registries, using anticonvulsants as a case study.

Methods: Pregnancies in women with epilepsy were identified and first trimester anticonvulsant exposure was determined. Major congenital malformations in the offspring were identified and verified. The risk of major congenital malformations following exposure to a range of anticonvulsants was calculated and compared to those reported by the UK Epilepsy and Pregnancy Register. The ability to identify a known teratogenic association using GPRD data was also assessed. An algorithm was created to identify and classify different types of pregnancy loss in an automated manner.

Results: The risks of a pregnancy outcome with a major congenital malformation following first trimester anticonvulsant exposures, were found to be similar in the GPRD to those of the UK Register. The number of exposures to individual products in the GPRD was often small and therefore lacked statistical power. It was, however, possible to identify a known teratogenic association using data from the GPRD. Verification of the algorithm developed to classify pregnancy losses demonstrated that, although not perfect, it would be a beneficial tool when using the GPRD for drug safety in pregnancy research.

Conclusion: It is unlikely a single data source or study design will be sufficient for monitoring all aspects of the safety of medicine use during pregnancy. The GPRD has the potential to make a valuable contribution to this field of research and could play an important role in complementing the work of other surveillance systems.

Table of contents

	Page
List of tables	6
List of figures	8
List of abbreviations	9
Acknowledgements	11
Introductory statement	12
1. Introduction	13
2. Alternative data sources to pregnancy exposure registries	20
2.1 An overview of pregnancy exposure registries	21
2.2 Literature search to identify alternative data sources	26
2.3 Discussion of alternative data sources	40
2.4 References	46
3. Aims and objectives	51
4. The General Practice Research Database (GPRD)	55
4.1 An overview of the General Practice Research Database	56
4.2 Features of the GPRD specific to drug safety in pregnancy research	62
4.3 Overview of drug safety in pregnancy research using the GPRD to date	66
4.4 References	70
5. Anticonvulsants as a case study	72
5.1 Treatment of epilepsy	73
5.2 Epilepsy and pregnancy	73
5.3 Anticonvulsant exposure and pregnancy outcomes	74
5.4 Pregnancy exposure registries for anticonvulsants	76
5.5 References	80
6. Identifying major congenital malformations in the UK General Practice Research Database	83
6.1 <i>Introduction</i>	84
6.2 <i>Methods</i>	84
6.2.1 Data source	84

6.2.2	Study population	85
6.2.3	Identification of major congenital malformations	86
6.2.4	Verification of congenital malformations	87
6.2.5	Analyses	87
6.3	<i>Results</i>	88
6.3.1	Verifying the presence of a congenital malformation	88
6.3.2	Classifying as a major or minor congenital malformation	92
6.4	<i>Discussion</i>	94
6.5	<i>Conclusion</i>	96
6.6	<i>References</i>	97
7.	Comparing the General Practice Research Database and the UK Epilepsy and Pregnancy Register as tools for postmarketing teratogen surveillance	98
7.1	<i>Introduction</i>	99
7.2	<i>Methods</i>	100
7.2.1	Data sources	100
7.2.2	General Practice Research Database maternal study population	101
7.2.3	Identifying pregnancy outcomes	102
7.2.4	Linking mother-baby pairs	102
7.2.5	Identifying major congenital malformations in live births	103
7.2.6	Identifying major congenital malformations in pregnancy losses	104
7.2.7	Determining first trimester exposure	104
7.2.8	Terminations of pregnancy	105
7.2.9	Analyses	105
7.3	<i>Results</i>	106
7.3.1	Study cohort	106
7.3.2	Major congenital malformations	108
7.3.3	Prevalence of major congenital malformations	110
7.3.4	Monotherapy exposures	113
7.3.5	Valproate and spina bifida	113
7.4	<i>Discussion</i>	115
7.4.1	Sample size	115
7.4.2	Cohort identification and disease status	115
7.4.3	Outcome assessment	116
7.4.4	Exposure assessment	117
7.4.5	Differences in risk assessment	118
7.4.6	Follow-up	118
7.4.7	Confounders	119
7.4.8	Comparison of data sources	119
7.5	<i>Conclusion</i>	119
7.6	<i>References</i>	120

8.	Recording of pregnancy losses on the General Practice Research Database	123
8.1	<i>Introduction</i>	124
8.2	<i>Methods</i>	125
8.2.1	Data source	125
8.2.2	Identification of pregnancy related medical codes	125
8.2.3	Study population	126
8.2.4	Identification and classification of pregnancy losses	126
8.2.5	Review of free text comments	127
8.2.6	Data analysis	127
8.3	<i>Results</i>	128
8.3.1	Identification of pregnancy losses	128
8.3.2	Categorisation of pregnancy losses	128
8.3.3	Verification of pregnancy losses	130
8.4	<i>Discussion</i>	132
8.5	<i>Conclusion</i>	135
8.6	<i>References</i>	136
9.	General discussion	137
9.1	Overall discussion of study results	138
9.2	Current developments	149
9.3	Overall conclusions	151
9.4	References	153
10.	Appendix I - Data resources for investigating drug exposure during pregnancy and associated outcomes: The GPRD as an alternative to pregnancy registries	154

List of tables

		Page
Table 2.1	Rationale for excluding articles identified via the literature search	29
Table 2.2	Summary of the data sources identified to evaluate the safety of medicine use during pregnancy	30
Table 5.1	Categories used by the US Food and Drug Administration to summarise the safety of medicines used during pregnancy based on the available evidence	79
Table 6.1	Percentage of congenital malformations that could be verified and, of those, the percentage that could be classified as major or minor	91
Table 6.2	The number of congenital malformations that could be verified, and of those the number that could be classified as major, by malformation class	93
Table 7.1	Population characteristics for mother-baby pairs where offspring were registered on the General Practice Research Database	108
Table 7.2	Number of malformations identified meeting the major congenital malformation inclusion criteria	109
Table 7.3	The age by which malformations in live-born infants were recorded in the General Practice Research Database	111
Table 7.4	Prevalence of major congenital malformations following different types of first-trimester antiepileptic drug exposure	112
Table 7.5	Prevalence of major congenital malformations following different monotherapy first-trimester antiepileptic drug exposures with carbamazepine as the reference category and then with the unexposed as the reference category	114
Table 8.1	Breakdown of the number of different types of pregnancy loss identified	130

Table 8.2	Comparison of algorithm endpoints with those derived from review of the free text comments	131
-----------	--	-----

List of figures

		Page
Figure 2.1	Human embryonic development showing sensitive periods	22
Figure 2.2	Identification of articles in the literature	28
Figure 4.1	A pregnant patient's medical record as viewed in the browser	61
Figure 6.1	Methods used for verifying malformations	88
Figure 6.2	Percentage of congenital malformations that could be verified using photocopied records and of those the percentage that could be classified as major or minor	89
Figure 6.3	Percentage of congenital malformations that could be verified using free text and of those the percentage that could be classified as major or minor	90
Figure 7.1	Identifying eligible pregnancies and mother-baby pairs	107
Figure 8.1	The number of pregnancy losses identified at each endpoint of the algorithm	129

List of abbreviations

AED	Antiepileptic drug
BMI	Body mass index
CI	Confidence interval
EDD	Estimated date of delivery
ENCePP	European Network of Centres for Pharmacoepidemiology and Pharmacovigilance
EUROCAT	European Concerted Action on Congenital Anomalies and Twins
FDA	Food and Drug Administration
GP	General practitioner
GPRD	General Practice Research Database
HES	Hospital Episode Statistics
ICBDSR	International Clearinghouse for Birth Defects Surveillance and Research
ISAC	Independent Scientific Advisory Committee to the GPRD
LCENS	Left censor date
LMP	Last menstrual period
MACDP	Metropolitan Atlanta Congenital Defects Program
MCM	Major congenital malformation
MHRA	Medicines and Healthcare products Regulatory Authority
MREC	Multi-center Research Ethics Committee
NHS	National Health Service
NTD	Neural tube defect
OXMIS	Oxford Medical Information System
PPV	Positive predictive value
RR	Relative risk
RCENS	Right censor date
SD	Standard deviation
SSRI	Selective serotonin reuptake inhibitor

TIS	Teratology Information Services
TOP	Termination of pregnancy
UK	United Kingdom
US	United States
UTS	Up-to-standard

Acknowledgements

First of all I must thank my supervisors Professor Corinne de Vries and Dr John Weil without whose continual guidance I would not be where I am today. Over the last five years they have always had my best interests in mind and the time, expertise and advice they have given me has been invaluable. I would also like to thank Dr Marianne Cunningham, Julia Snowball, Sayantani Ray, Deepak Sriramulu and Chanchal Bains for all their individual contributions to the work that has been presented in this thesis. Thanks must also go to my family and friends who have provided the sometimes much needed encouragement to enable me to complete this PhD.

Introductory statement

This thesis reports on the potential of using the General Practice Research Database as an alternative to pregnancy exposure registries for monitoring the safety of medicine use during pregnancy. This thesis contains published and unpublished research. Chapters 6 and 7 have been published as peer reviewed original research; Chapter 8 will be submitted for publication. The publications are the product of a joint collaboration between the University of Bath and GlaxoSmithKline. The study was largely funded by GlaxoSmithKline.

For the work presented in Chapters 6 and 7, I led on the protocol development with valuable input from individuals at both GlaxoSmithKline and the University of Bath. I was responsible for identifying all anticonvulsant product codes and medical codes relating to pregnancy, epilepsy and congenital malformations. The data extraction from the GPRD was performed by Sayantani Ray at GlaxoSmithKline. All verification of congenital malformations and analyses were carried out by me at the University of Bath.

For the work presented in Chapter 8, Corinne de Vries and Julia Snowball at the University of Bath identified all pregnancy codes and Julia Snowball wrote the program for the termination of pregnancy algorithm. I was involved in reviewing all medical codes, manually reviewing patients' medical records and comparing a sample of pregnancy losses identified and classified using the algorithm with those from the manual review. Amendments to the algorithm were subsequently made based on the manual review comparison carried out by myself and others at the University of Bath. I carried out the verification of a sample of algorithm endpoints by reviewing additional non-coded information recorded by GPs in the patients' medical records.

The inclusion of manuscripts that were written up for publication within this thesis does result in a small amount of repetition in the methods sections.

Chapter 1

Introduction

Before the 1960s, when it was discovered thalidomide was a major teratogen,⁽¹⁾ little consideration had been given to the lack of available information regarding the safety of medicines when used by women during pregnancy. The realisation that thalidomide, which had been specifically marketed to pregnant women, was responsible for causing serious birth defects in several thousand infants exposed to the drug *in utero*,⁽²⁾ alerted governments to the potential dangers associated with medicine use during pregnancy. Following this discovery, the need for tighter control and regulation of all aspects of the licensing of new medicinal products was recognised and, in particular, the need to evaluate the safety of medicine use during pregnancy.

The percentage of pregnancies exposed to medicine use has been reported to be anywhere between 27 and 99%⁽³⁻⁶⁾ depending on the country of study, the definition of 'medicine use' and whether the first trimester alone or the entire pregnancy period was considered. This high percentage consists of a combination of different forms of exposure. In some circumstances exposure will be unavoidable, owing to the potential risks to both mother and foetus of the underlying medical condition (e.g. epilepsy) which makes discontinuation of treatment inadvisable. There will also be exposures that occur as the result of the woman developing a new medical condition during pregnancy that requires treatment (e.g. an infection). In addition, many exposures to prescription medicines during pregnancy will occur inadvertently and these largely result from the fact that between 30-50% of pregnancies are unplanned.^(7, 8) Women can also be exposed to products that have been on the market for a long time, which are often available over-the-counter, without a prescription, where there can be a false perception that they are safe, even though their safety when used during pregnancy has not been specifically evaluated. Knowledge of a particular medicine's safety in relation to pregnancy, is important to provide women and healthcare professionals with sufficient information to enable them to make informed decisions about their treatment. In many cases this knowledge can also provide some level of reassurance following inadvertent exposures, helping to

avoid undue concern and potentially the unnecessary decision to have a termination of pregnancy.

There are many theories and methods that are applied in the general field of drug safety that cannot be applied to drug safety in pregnancy. Although animal studies, in some instances, can alert scientists to areas of potential concern, the teratogenic effects observed in animals cannot always be extrapolated to predict those in humans.^(9, 10) Pregnant women are often excluded from clinical trials and the number of inadvertent exposures is too small to provide any reliable evidence on potential increases in risk.⁽¹¹⁾ In addition, it is not possible to assume a 'class effect' as drugs within the same drug class do not always act in a uniform manner. Although data on a drug's pharmacology and toxicology can be available, the substantial gaps in our knowledge of teratogenic mechanisms mean that it is often of limited value.⁽¹²⁾ The field of drug safety in pregnancy also differs owing to the further complication of the foetus as an 'innocent bystander',⁽¹²⁾ a treatment that is required for the benefit of one individual has the potential to have an adverse effect on a second individual, who can expect only indirect benefits (i.e. a healthy mother). As a result of all these factors, evaluating the safety of a medicine when used during pregnancy is virtually impossible before it is granted a licence. Therefore much of the focus of the assessment of a product's safety has to be given to monitoring pregnancy outcomes in women who use the product during pregnancy, once it is on the market.

The last fifty years have seen the introduction of a number of methods to evaluate the safety of medicine use during pregnancy with the risk of major congenital malformations often being the main outcome of interest. Initially drug safety in pregnancy surveillance consisted of the collation of spontaneous reports of adverse events sent by healthcare professionals and patients to pharmaceutical companies.⁽¹³⁾ The late 1970s and early 1980s then saw the development of a number of case-control surveillance systems that recruited infants with and without congenital malformations, with the aim of evaluating

associations between medicine use and pregnancy outcomes.⁽¹⁴⁾ Some of these systems are still actively recruiting today⁽¹⁵⁾ and unlike spontaneous reports, the nature of the case-control study design means they are particularly suited to testing hypotheses that have been generated by other types of surveillance systems. At a similar time, the first teratology information services (TIS) were set up with the aim of identifying exposed pregnancies and following them up to capture pregnancy outcome data.⁽¹⁶⁾ Later, the role of the TIS was broadened to also act as a point of contact for women and healthcare professionals who needed information and advice regarding *in utero* drug exposures. In 1979 the European Concerted Action on Congenital Anomalies and Twins (EUROCAT) was set up. EUROCAT aimed to capture and combine European-wide registrations of congenital anomalies to identify evidence of clusters associated with environmental exposures.⁽¹⁷⁾ In addition to environmental exposures, when registering a congenital anomaly there is also the option of recording information on maternal drug exposure. In recent years EUROCAT has begun to utilise this information to evaluate the safety of some medicines when used during pregnancy.⁽¹⁸⁾

In 1984 the pharmaceutical company Burroughs Wellcome Co. established the first pregnancy exposure registry as a tool to monitor the safety of acyclovir when used by pregnant women.⁽¹⁹⁾ Since then, pregnancy exposure registries have become one of the most commonly used methods for identifying pregnant women and pregnancy outcomes in order to evaluate the safety, largely of new medicines, when used during pregnancy.

Despite the development of spontaneous reporting systems, teratology information services, case-control surveillance systems and pregnancy exposure registries, fifty years after the thalidomide disaster there are still a large number of medicines, both new and old, where little, if anything, is known about their safety when used by pregnant women. The introduction of risk management and risk minimisation plans and the United States Food and Drug Administration (FDA) 2007 Amendment Act⁽²⁰⁾ all put increasing emphasis on the requirement

for market authorisation holders to be proactive in evaluating the safety of their products, when used during pregnancy.⁽²¹⁾ Given the current increase in the electronic capture of healthcare data for administrative, clinical audit and research purposes and given the limitations of some existing data sources and methods, increasingly, electronic healthcare and claims data are being used for drug safety studies – including drug safety in pregnancy. This has been done both for signal strengthening or hypothesis testing and for hypothesis generating studies. As these increase in popularity and given the methodological challenges specific to drug safety in pregnancy research, there is a need to evaluate these data sources individually to determine their potential as new and complementary data sources in this area.

This PhD aims to evaluate the potential of the United Kingdom's General Practice Research Database (GPRD) to be used as an alternative or complement to pregnancy registries in monitoring the safety of medicines when used by pregnant women. The next chapter will discuss the strengths and limitations of pregnancy exposure registries, before introducing the reader to additional data sources that are currently available and used in this area of research. The work presented in this chapter builds on that published in 2008,⁽²²⁾ before I registered to study for a PhD. This is followed by an outline of the specific aims and objectives of the PhD (Chapter 3). Chapter 4 then provides information on the GPRD, discusses features of the database that are specific to monitoring the safety of medicine use during pregnancy and summarises key research in this field using the GPRD to date. The focus of the thesis is exposure to anticonvulsants; the fifth chapter explains the rationale behind this choice. The methods and results of a study that verified congenital malformations identified in the GPRD and evaluated the sensitivity and added value of photocopied medical records and free text are discussed in Chapter 6. Chapter 7 presents the findings of a study that compared the rates of major congenital malformations in the GPRD with those reported by the UK Epilepsy and Pregnancy Register. This chapter also reports on the potential of the GPRD, when I attempted to identify a known teratogenic association. Chapter 8 outlines the importance of including

pregnancy losses when assessing the safety of medicines used during pregnancy and reports on work carried out to identify and categorise pregnancy losses in an automated manner in the GPRD. In the final chapter I outline how I have met each of the objectives set out in Chapter 3 and discuss the findings and their implications. Then, based on all the evidence presented in this thesis, I will draw final conclusions on the potential of the GPRD in the field of drug safety in pregnancy research and the areas where I believe it has been demonstrated it can make a valuable contribution.

References

1. McBride W. Thalidomide and congenital abnormalities. *Lancet*. 1961;2:1358.
2. Lenz W. A short history of thalidomide embryopathy. *Teratology*. 1988;38(3):203-15.
3. Daw JR, Hanley GE, Greyson DL, et al. Prescription drug use during pregnancy in developed countries: a systematic review. *Pharmacoepidemiol Drug Saf*. 2011;20(9):895-902.
4. Cleary BJ, Butt H, Strawbridge JD, et al. Medication use in early pregnancy-prevalence and determinants of use in a prospective cohort of women. *Pharmacoepidemiol Drug Saf*. 2010;19(4):408-17.
5. Lacroix I, Damase-Michel C, Lapeyre-Mestre M, et al. Prescription of drugs during pregnancy in France. *Lancet*. 2000;356(9243):1735-6.
6. Garriguet D. Medication use among pregnant women. *Health Reports*. 2006;17(2):9-18.
7. Henshaw SK. Unintended pregnancy in the United States. *Fam Plann Perspect*. 1998;30(1):24-9.
8. Lakha F, Glasier A. Unintended pregnancy and use of emergency contraception among a large cohort of women attending for antenatal care or abortion in Scotland. *Lancet*. 2006;368(9549):1782-7.
9. Hoar RM. Developmental toxicity: extrapolation across species. *J Am Coll Toxicol*. 1999;14(1):11-20.
10. Koren G, Pastuszak A, Ito S. Drugs in pregnancy. *N Engl J Med*. 1998;338(16):1128-37.
11. Shields KE, Wiholm B, Hostetler LS, et al. Monitoring outcomes of pregnancy following drug exposure: a company-based pregnancy registry program. *Drug Saf*. 2004;27(6):353-67.
12. Mitchell A. Studies of drug-induced birth defects. In: Strom BL, editor. *Pharmacoepidemiology*: John Wiley and Sons Ltd; 2005. p. 501-14.
13. Cunnington M, Messenheimer J. Pregnancy Registries: strengths, weaknesses, and bias interpretation of pregnancy registry data. *Int Rev Neurobiol*. 2008;83:283-304.

14. Werler MM, Louik C, Mitchell AA. Case-control studies for identifying novel teratogens. *Am J Med Genet C Semin Med Genet*. 2011;157(3):201-8.
15. Louik C, Schatz M, Hernández-Díaz S, et al. Asthma in pregnancy and its pharmacologic treatment. *Ann Allergy Asthma Immunol*. 2010;105(2):110-7.
16. Leen-Mitchell M, Martinez L, Gallegos S, et al. Mini-review: History of organized teratology information services in North America. *Teratology*. 2000;61(4):314-7.
17. Dolk H. EUROCAT: 25 years of European surveillance of congenital anomalies. *Arch Dis Child Fetal Neonatal Ed*. 2005;90:F355-F8.
18. Dolk H, Jentink J, Loane M, et al. Does lamotrigine use in pregnancy increase orofacial cleft risk relative to other malformations? *Neurology*. 2008;71:714-22.
19. Andrews EB, Yankaskas BC, Cordero JE, et al. Acyclovir in pregnancy registry: six years' experience. The Acyclovir in Pregnancy Registry Advisory Committee. *Obstet Gynecol*. 1992;79(1):7-13.
20. HR 3580 FDA Amendments Act, 2007, Sec 905; D.i.II.Bb.
21. Wyszynski DF. Pregnancy exposure registries: Academic opportunities and industry responsibility. *Birth Defects Res A Clin Mol Teratol*. 2009;85:93-101.
22. Charlton RA, Cunnington MC, de Vries CS, et al. Data resources for investigating drug exposure during pregnancy and associated outcomes: the General Practice Research Database (GPRD) as an alternative to pregnancy registries. *Drug Saf*. 2008;31(1):39-51.

Chapter 2

Alternative data sources to pregnancy exposure registries

2.1 An overview of pregnancy exposure registries

Pregnancy exposure registries are essentially prospective observational studies that follow women up from the time of enrolment in the registry until a short period after pregnancy outcome. They are created with the aim of detecting major teratogenicity, that is, where a large proportion (e.g. 30-40%) of those pregnancies exposed to a particular drug are adversely affected.⁽¹⁾ Pregnancy exposure registries can be set up either by pharmaceutical companies, academic groups or research groups, they can be international or country specific and they can focus on a single drug, a drug class or a disease. The European Medicines Agency and the FDA recommend pharmaceutical companies consider developing a pregnancy exposure registry for products that may be used during pregnancy to treat new or chronic conditions and for products frequently used by women of childbearing age where the likelihood of inadvertent exposure during pregnancy is high.^(2, 3)

Pregnancy exposure registry methods

The precise methodology used can vary slightly between registries but in general, women can enrol either directly themselves or via one of their healthcare providers (GP, midwife, epilepsy nurse etc.). Enrolment should ideally be before any prenatal screening has taken place and before the pregnancy outcome is known in order to avoid selection bias towards more severe outcomes. At the time of enrolment, informed consent is obtained and information is collected on some or all of the following: general demographics, use and timing of prescription and over-the-counter medicines, disease status (e.g. number / type of epilepsy seizures), potential confounding factors including smoking status, alcohol consumption and folic acid exposure. Given our knowledge of the different stages of foetal development, pregnancy registries have tended to focus their analysis on pregnancies where drug exposure occurred during the

first trimester of pregnancy as this is the time period of greatest susceptibility in terms of the risk of major congenital malformations^a (Figure 2.1).

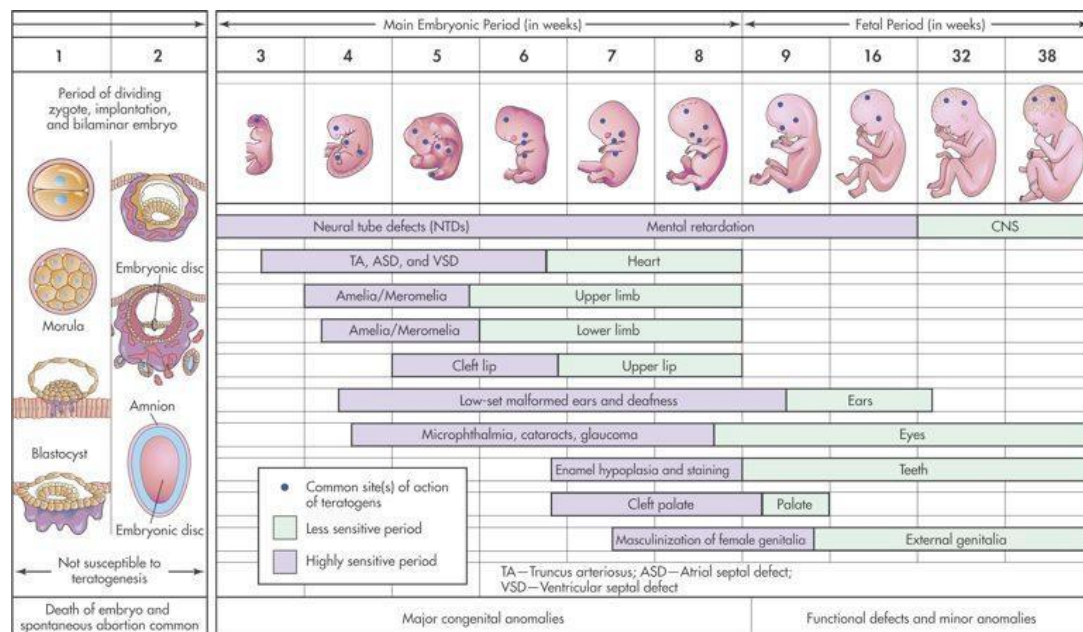


Figure 2.1 Human embryonic development showing sensitive periods (From Moore, 1988⁽⁴⁾ with permission)

Follow up information on the pregnancy outcome and the presence or absence of a congenital malformation is collected, shortly after the expected date of delivery, by a GP or patient questionnaire or telephone call. Live births, stillbirths, induced terminations and spontaneous abortions are captured by registries although the number of spontaneous pregnancy losses captured may be relatively low depending on the week's gestation at which women enrol. The primary endpoint of a pregnancy registry is an estimate of the overall risk of all major congenital malformations⁽⁵⁾ with the aim of providing data based on exposures in humans that is clinically relevant and can be used to inform healthcare professionals and patients.⁽⁶⁾ In addition to collecting information on congenital malformations, some registries have chosen to extend the length of infant follow-up in order to evaluate any evidence of an association between maternal drug exposure and developmental delay in the offspring.⁽⁵⁾

^a Major congenital malformations will be discussed in more detail in Chapters 6 and 7 but they are broadly defined as abnormalities present at birth that are of surgical, medical or cosmetic importance

To reduce the likelihood of selection bias, analysis of data collected by pregnancy exposure registries tends to focus on those pregnancies that were prospectively enrolled before any prenatal screening or knowledge of the pregnancy outcome has occurred. Pregnancies reported to registries retrospectively, following the diagnosis of a major congenital malformation, are still reviewed and analysed because they may help to identify multiple cases of the same defect type, which would require further investigation.⁽⁶⁾

In addition to the main aim of identifying major teratogenicity, pregnancy exposure registries can also act as hypothesis-generating studies by detecting adverse pregnancy outcomes that may warrant further investigation. To do this many pregnancy registries have adopted the 'rule of 3' where review is thought warranted if the registry observes 3 or more reports to be of a particular defect following the same exposure. The 'rule of 3' is based on the rationale that in a registry with fewer than 600 exposures, the likelihood of observing 3 of the same specific birth defect when it normally occurs with a rate of less than 1/700 is unlikely to be by chance alone.⁽⁷⁾

Limitations of pregnancy exposure registries

Although pregnancy registries have several strengths over other surveillance methods it is widely recognised that they also have a number of limitations.

Enrolment

Low levels of enrolment are commonly found to hinder pregnancy exposure registries. The European Committee for Medical Products for Human Use considers 1000 exposures to be representative of widespread market exposure,⁽⁸⁾ yet the pharmaceutical company GlaxoSmithKline has sponsored five international registries, none of which managed to enrol 1000 pregnancies with informative outcomes during their first ten years of data collection.⁽⁹⁾ Attempts to raise awareness and encourage enrolment are often hampered by the lack of knowledge regarding the safety of the product being monitored, making it difficult to decide on how to communicate the message and the need to ensure

any promotional material does not appear to encourage use of the product or give a false impression of safety.^(6, 10)

The voluntary nature of enrolment can result in selection bias if women opting to enrol differ from those who do not, in terms of factors associated with the underlying risk of the outcome being studied.⁽¹¹⁾ For example, women choosing to enrol into a registry may be more health conscious and more likely to follow advice in relation to the potential benefits of pre-conceptional folic acid, smoking cessation and reducing alcohol intake during pregnancy than those who do not. In addition to selection bias resulting from enrolment by the women themselves, registries may also suffer from referral bias with healthcare professionals being more or less likely to enrol women with a particular disease severity or those exposed to a particular type of treatment. To my knowledge thus far no comparisons have been published comparing the population characteristics and disease severity for individuals enrolled in a pregnancy registry with those from a representative sample of individuals who would be eligible to enrol.

Loss to follow-up

Pregnancy exposure registries often suffer from loss to follow-up. This has been reported to be as low as 8.1% in the UK Epilepsy and Pregnancy Register⁽¹²⁾ and as high as 35.8% in the Bupropion Pregnancy Registry.⁽¹³⁾ In 2004, in an attempt to reduce loss to follow-up, three pregnancy registries trialled the introduction of a stipend for healthcare professionals who reported follow-up pregnancy outcome data to the registry. Analysis of loss to follow-up rates before and after this introduction found the incentive of a stipend, to reimburse healthcare professionals for the time taken to report follow-up pregnancy outcome data, did not significantly reduce the proportion of pregnancies lost to follow-up.⁽¹⁴⁾

Statistical power

A combination of low enrolment, loss to follow-up and a low frequency of the exposure and outcome of interest can limit the statistical power and validity of pregnancy exposure registries. At best, pregnancy registries are often only

powered to detect major teratogens and evaluate the risk of all major congenital malformations combined. There may, however, be instances where a registry generates a signal relating to an increased risk of a particular defect type.⁽¹⁵⁾ In these instances, although data from other pregnancy registries monitoring the same exposure can be analysed in an attempt to confirm or refute the possible association, it is likely that they too will lack statistical power and therefore additional data sources will be required to investigate further.

Information on potential confounders

When sample sizes are small, the inclusion of too many confounding variables can make any statistical models of risk assessment unstable but as more individuals are enrolled the number of confounding variables considered can potentially be increased.⁽¹⁶⁾ Pregnancy exposure registries, however, require primary data collection, which can be both costly and time consuming. This can often mean that less information on potential confounding variables is requested so as not to dissuade pregnant women and healthcare professionals from choosing to enrol.⁽¹⁷⁾ For the identification of a high-risk teratogen a lack of this information, although restrictive, is unlikely to dramatically alter the risk estimates.⁽¹⁾

Comparator group

The selection of a suitable comparator group when evaluating data from pregnancy exposure registries is challenging, especially when there is a possibility that the medical condition that the treatment is for may itself be associated with the outcome of interest (e.g. diabetes, epilepsy).⁽¹¹⁾ There are many possible comparator groups that can be used and the most appropriate will depend on the question being asked and the exposure and outcome of interest. Some analyses carried out by registries involve making comparisons with population-based birth defect surveillance systems such as the Metropolitan Atlanta Congenital Defects Program (MACDP),⁽⁷⁾ some make comparisons with other monotherapy exposures that have been collected via the registry, some registries enrol women who have the disease but were not treated during

pregnancy,⁽¹²⁾ some enrol their own unexposed comparator group such as family or friends of the exposed woman⁽¹⁸⁾ and some make multiple comparisons using a combination of the comparator groups mentioned. It could be argued however, that given the aim, to identify major teratogenicity, no formal comparator group is needed and instead comparison with background prevalence should be sufficient.^(1, 19)

2.2 Literature search to identify alternative data sources to pregnancy exposure registries

Pregnancy exposure registries have been successful in both providing reassurance that certain products are not major teratogens⁽²⁰⁾ and in generating signals of potential teratogenicity that require further investigation.⁽¹⁵⁾ Their limitations, however, along with the acceptance that a single data source is unlikely to be sufficient to provide all the answers, have led researchers to look for alternative and complementary sources of data for evaluating prenatal drug exposures.

One alternative type of data source that is becoming the focus of much research is that of electronic healthcare databases. Electronic databases are increasingly being used to manage medical insurance claims and patient medical records and this has resulted in an ever-growing volume of healthcare data being available for pharmacoepidemiology research. The initial signal that suggested a possible association between first trimester exposure to paroxetine (a selective serotonin reuptake inhibitor (SSRI)) and an increased risk of major congenital malformations and cardiovascular defects resulted from a study based on electronically recorded healthcare claims data from the United States.⁽²¹⁾ Given that there was no pregnancy exposure registry set up for paroxetine this potential association could have otherwise gone undetected. Following the initial study a number of other studies were conducted using a range of different data sources and epidemiological study designs in order to try to confirm or refute the association.⁽²²⁻²⁶⁾ The findings of these studies ultimately resulted in changes being made to the product label.⁽²⁷⁾

Additional sources of information on drug exposures during pregnancy and pregnancy outcomes have the ability to complement pregnancy exposure registries in a number of ways. The remainder of this chapter reports on a literature review carried out to identify additional data sources that are currently being used to monitor the safety of medicine use during pregnancy. This review builds on the review that was published in January 2008 (Appendix I).⁽⁹⁾

Methods

A review of the literature was conducted to identify papers (excluding conference abstracts) reporting on the safety of medicine use during pregnancy that had used a data source which had systematic data collection. In PubMed papers were identified based on the following search: (('Pregnancy'[Mesh] OR 'Congenital Abnormalities'[Mesh] OR 'Teratogens'[Mesh]) AND ('Product Surveillance, Postmarketing'[Mesh])), whilst in Embase papers were identified based on (('Pregnancy' OR 'Pregnancy outcome' OR 'Pregnancy termination' OR 'Congenital disorder' OR 'Congenital malformation' OR 'Birth defects' OR 'Teratogenic agent' OR 'Teratogenicity') AND ('Postmarketing surveillance' OR 'Drug surveillance program')) and were restricted to papers reporting on studies in Humans. All papers were restricted to those published in English between 1 January 2000 and 30 November 2011. In addition to searching the literature, individuals who are specialists in the field of drug safety in pregnancy were consulted to ensure any additional data sources were captured.

Results

The literature searches identified 236 articles through PubMed and a further 381 articles via Embase. Of these 505 were excluded following review of the title and abstract and a further 30 were excluded following review of the full text (Figure 2.2).

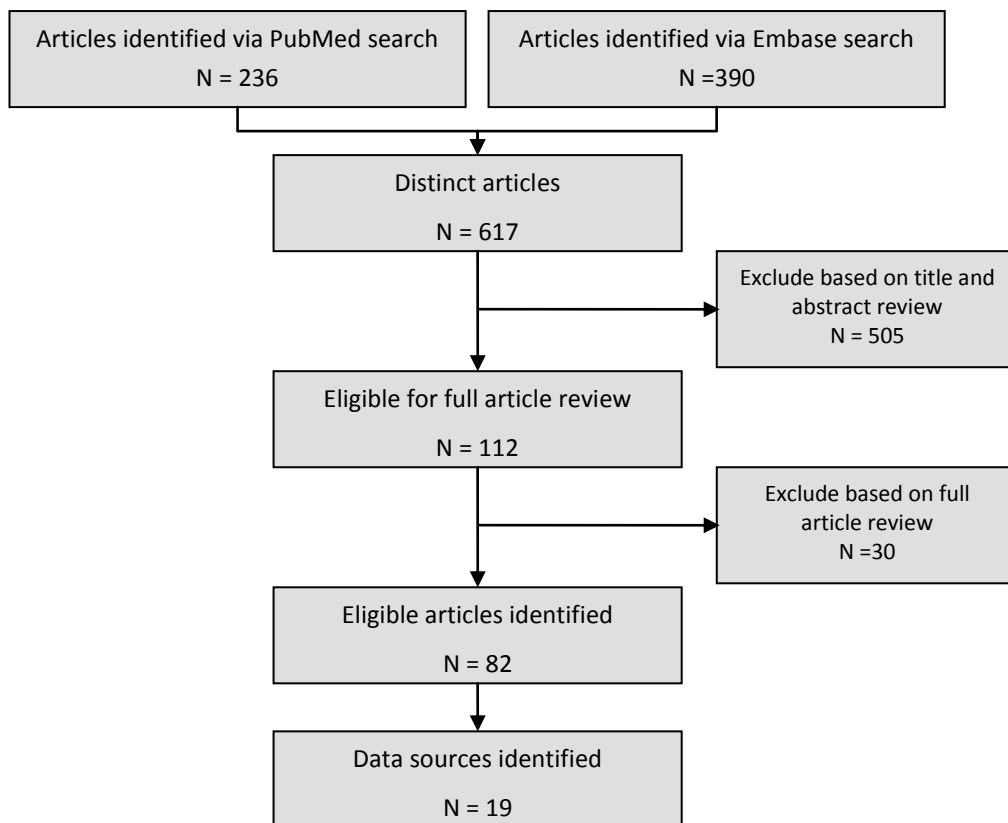


Figure 2.2 Identification of articles in the literature

Table 2.1 summarises the rationale for excluding the papers identified and excluded at this stage. A total of 82 articles were included. Overall the studies reported on used data from 19 different data sources. A further 6 data sources were identified and included as a result of my knowledge of the sources available and by contacting specialists in the field. Table 2.2 provides an overview of each of the 25 data sources identified. Where the papers identified via the literature search did not have sufficient information to complete all the fields in the table, additional papers reporting on those sources were identified. Where information was still missing, the authors of the papers were contacted.

Table 2.1 Summary of the rationale for those articles excluded

Reason for exclusion	Number of articles excluded
Pregnancy exposure registries	23
Teratology information centres	17
Field studies with one-off manual data collection	16
Meta-analyses	7
Spontaneous/case reports	11
Environmental or occupational exposures	19
Alcohol or illicit drug use exposures	6
Overview of teratogenicity in general or pregnancy exposure registries	47
Comments or letters to the editor	97
Review papers	154 ^b
Other (e.g. product surveillance in general - not specifically pregnancy, reviews of medical conditions during pregnancy)	138 ^b

^b These categories are large but they mainly come from the Embase search where the search strategy is not as refined as in PubMed and this results in a large number of unrelated publications being identified

Table 2.2 Summary of the data sources identified to evaluate the safety of medicine use during pregnancy

Key: Dark text represents those variables captured by the data source and light text represents those variables that are not available.

Name of data source	Time period of data collection	Population covered	Source of exposure information	Types of pregnancy outcome captured	Source of outcome information	Additional risk information (all capture maternal age)
Population-based record linkage surveillance systems						
Swedish Medical Birth Register ^(22, 28)	Medical birth register since 1973, including drug use since July 1994 Prescribed drug register since 2005	Country Sweden Population-based – Yes ~98% of all deliveries Sample size ~110,000 births per year	Maternal self reporting at first antenatal interview and copies of antenatal care records are reviewed Prescribed drug register of filled prescriptions since 2005	<ul style="list-style-type: none"> • Live births • Stillbirths • Spontaneous losses • Elective terminations 	Identified from the Register of Birth Defects and the Patient Register – data recorded by a paediatrician Opportunity for medical record review - Yes	<ul style="list-style-type: none"> • Smoking status • Alcohol consumption • Body mass index • Socioeconomic status • Maternal diagnoses • Co-prescribing • Folic acid <ul style="list-style-type: none"> - if reported • Over-the-counter medicines <ul style="list-style-type: none"> - if reported • Reproductive history
Norwegian Medical Birth Register ^(29, 30)	Medical birth registry of Norway since 1967, including drug use since 1998 Norwegian Prescription database since 2004	Country Norway Population-based – Yes Compulsory reporting of all births and late abortions from 12 weeks gestation Sample size ~60,000 births per year	Recorded during antenatal visits to GP, midwife and obstetrician. Potential to use prescribed drug register of filled prescriptions since 2004	<ul style="list-style-type: none"> • Live births • Stillbirths • Spontaneous losses • Elective terminations <ul style="list-style-type: none"> - from 12 weeks gestation 	Recorded by physicians and midwives Opportunity for medical record review - Yes	<ul style="list-style-type: none"> • Smoking status – since 1998 • Alcohol consumption • Body mass index • Socioeconomic status • Maternal diagnoses • Co-prescribing • Folic acid <ul style="list-style-type: none"> - Since 1998 • Over-the-counter medicines <ul style="list-style-type: none"> - if the GP is aware • Reproductive history

Finnish linked national health registers ^(31, 32)	<p>Medical birth register since 1987</p> <p>Register on induced abortions since 1977</p> <p>Register of reimbursement drugs since 1994</p>	<p>Country Finland</p> <p>Population-based – Yes Compulsory reporting of all deliveries and elective terminations</p> <p>Sample size ~58,000 deliveries and ~ 10,500 elective terminations per year</p>	<p>Information on reimbursed purchases of prescription medicines from the Register of Reimbursement Drugs</p>	<ul style="list-style-type: none"> • Live births • Stillbirths • Spontaneous losses • Elective terminations <p>Will have spontaneous losses treated in hospital and primary care from 2011</p>	<p>Identified from the register of congenital malformations – data recorded by hospital personnel.</p> <p>Opportunity for medical record review - Yes</p>	<ul style="list-style-type: none"> • Smoking status • Alcohol consumption • Body mass index • Socioeconomic status • Maternal diagnoses - chronic • Co-prescribing • Folic acid – high dose only • Over-the-counter medicines • Reproductive history
Danish National Patient Registry ⁽³³⁻³⁵⁾	<p>Danish National Patient Registry since 1996</p> <p>Prescription data from 1995 but only available since 2003</p>	<p>Country Denmark</p> <p>Population-based – Yes Compulsory reporting of all births</p> <p>Sample size ~50,000 deliveries per year</p>	<p>Filled prescription data from the Registry of Medicinal Product Statistics since 2003</p> <p>Previously would have been self reported via maternal interview</p>	<ul style="list-style-type: none"> • Live births • Stillbirths • Spontaneous losses • Elective terminations 	<p>Routinely recorded inpatient and outpatient data recorded by paediatrician</p> <p>Opportunity for medical record review - Yes</p>	<ul style="list-style-type: none"> • Smoking status • Alcohol consumption • Body mass index • Socioeconomic status • Maternal diagnoses - hospital diagnoses only • Co-prescribing • Folic acid – high dose only • Over-the-counter medicines • Reproductive history
The North Jutland Pharmaco-Epidemiological Prescription Database with linked registries ⁽³⁶⁾	<p>Prescription database since 1991</p> <p>Danish National Patient Registry since 1996</p>	<p>Country Denmark</p> <p>Population-based – Yes County of North Jutland - compulsory reporting of all births</p> <p>Sample size ~6,000 deliveries a year</p>	<p>Dispensed prescription data used to secure reimbursement from the Health Service to the pharmacies</p>	<ul style="list-style-type: none"> • Live births • Stillbirths • Spontaneous losses • Elective terminations 	<p>County hospital Discharge Register – discharge diagnoses recorded by paediatrician</p> <p>Opportunity for medical record review - Yes</p>	<ul style="list-style-type: none"> • Smoking status • Alcohol consumption • Body mass index • Socioeconomic status • Maternal diagnoses - hospital diagnoses only • Co-prescribing • Folic acid – high dose only • Over-the-counter medicines • Reproductive history

Saskatchewan population registries ^(37, 38)	Hospital data from 1970 Prescription data from 1975	Country Canada Population-based – Yes Covers >90% of the Canadian province Sample size ~11,400 deliveries per year	Dispensed prescriptions on the Outpatient Prescription Drug Database	<ul style="list-style-type: none"> • Live births • Stillbirths • Spontaneous losses • Elective terminations 	Identified from the Hospital Services Database – data recorded electronically by physician Opportunity for medical record review - Yes	<ul style="list-style-type: none"> • Smoking status • Alcohol consumption • Body mass index • Socioeconomic status • Maternal diagnoses • Co-prescribing • Folic acid • Over-the-counter medicines • Reproductive history
Taiwan National Health Insurance Research Dataset linked to the Birth Certificate Registry ⁽³⁹⁾	Since 1996	Country Taiwan Population-based – Yes ~98% of the Taiwan population Sample size ~200,000 births per year	Dispensed prescription data recorded in the National Health Insurance Research Dataset	<ul style="list-style-type: none"> • Live births • Stillbirths • Spontaneous losses • Elective terminations 	Identified from medical claims recorded in the National Health Insurance Research Dataset Opportunity for medical record review - No	<ul style="list-style-type: none"> • Smoking status • Alcohol consumption • Body mass index • Socioeconomic status - maternal education only • Maternal diagnoses • Co-prescribing • Folic acid • Over-the-counter medicines • Reproductive history
Western Australia population-based Data Linkage System ^(40, 41)	Since 2002 Birth defect registry since 1980	Country Australia Population-based – Yes All pregnancies in Western Australia Sample size ~ 40,000 pregnancies a year	Dispensed prescriptions. Covers those issued in community and private hospitals and from 2004 public hospitals that are subsidised ~80% of all prescriptions	<ul style="list-style-type: none"> • Live births • Stillbirths • Spontaneous losses • Elective terminations <p>Looking into linking elective terminations with the birth defect registry</p>	Notifications received from paediatricians, obstetricians, cytogenetics, ultrasound, genetic counselling departments to the Birth Defects Registry of western Australia.	<ul style="list-style-type: none"> • Smoking status • Alcohol consumption • Body mass index • Socioeconomic status • Maternal diagnoses • Co-prescribing • Folic acid • Over-the-counter medicines • Reproductive history

Region Emilia-Romagna (RER) Database ⁽⁴²⁾	Since 2000	Country Italy Population-based – Yes ~99% of pregnancies in Region Emilia-Romagna Sample size ~ 33,000 pregnancies a year	Reimbursed prescription data (~70% of medicines can be reimbursed)	<ul style="list-style-type: none"> • Live births • Stillbirths • Spontaneous losses • Elective terminations 	Hospital assistance at birth records, hospital discharge records and links to Congenital anomaly register Opportunity for medical record review - No	<ul style="list-style-type: none"> • Smoking status • Alcohol consumption • Body mass index • Socioeconomic status • Maternal diagnoses • Co-prescribing • Folic acid • Over-the-counter medicines • Reproductive history
Healthcare databases Medical record databases						
General Practice Research Database ^(43, 44)	Since 1987	Country United Kingdom Population-based – Yes ~8% sample of the UK population Sample size ~80,000 pregnancies per year	Prescriptions issued by GPs and recorded in medical records	<ul style="list-style-type: none"> • Live births • Stillbirths • Spontaneous losses • Elective terminations 	Diagnoses recorded in medical records by GPs Opportunity for medical record review - Yes	<ul style="list-style-type: none"> • Smoking status • Alcohol consumption • Body mass index • Socioeconomic status • Maternal diagnoses • Co-prescribing • Folic acid – high dose only • Over-the-counter medicines • Reproductive history
The Health Improvement Network (THIN) ^(45, 46)	Since 2003	Country United Kingdom Population based – Yes ~6% sample of the UK population Sample size ~60,000 pregnancies per year	Prescriptions issued by GPs and recorded in medical records	<ul style="list-style-type: none"> • Live births • Stillbirths • Spontaneous losses • Elective terminations 	Diagnoses recorded in medical records by GPs Opportunity for medical record review - Yes	<ul style="list-style-type: none"> • Smoking status • Alcohol consumption • Body mass index • Socioeconomic status • Maternal diagnoses • Co-prescribing • Folic acid – high dose only • Over-the-counter medicines • Reproductive history

Administrative claims databases						
Tennessee Medicaid ^(47, 48)	Since 1985	Country United States Population-based – No - generally low income adults Sample size ~36,000 deliveries per year	Pharmacy claims data for dispensed prescriptions	<ul style="list-style-type: none"> • Live births • Stillbirths • Spontaneous losses • Elective terminations 	Identified from Medicaid inpatient, emergency department physician visit, hospital, discharge diagnoses records Also link to birth and foetal death certificates Opportunity for medical record review - Yes	<ul style="list-style-type: none"> • Smoking status • Alcohol consumption • Body mass index • Socioeconomic status • Maternal diagnoses - chronic • Co-prescribing • Folic acid • Over-the-counter medicines • Reproductive history
Kaiser Permanente ^(49, 50)	Since ~1995	Country United States Population-based – No Under-represents those at the extremes of household income Sample size ~30,000 deliveries per year	Pharmacy claims data for dispensed prescriptions	<ul style="list-style-type: none"> • Live births • Stillbirths • Spontaneous losses • Elective terminations 	Medical claims records Opportunity for medical record review - Yes	<ul style="list-style-type: none"> • Smoking status • Alcohol consumption • Body mass index • Socioeconomic status - maternal education only • Maternal diagnoses • Co-prescribing • Folic acid • Over-the-counter medicines • Reproductive history
United Healthcare ^(21, 51, 52)	Since 1990	Country United States Population-based – No ~2% of US population. 90% are employer groups,	Electronically recorded dispensed prescription data	<ul style="list-style-type: none"> • Live births • Stillbirths • Spontaneous losses • Elective terminations 	Medical claims records from inpatient, hospital, outpatient, emergency department, surgery centre and physician's	<ul style="list-style-type: none"> • Smoking status • Alcohol consumption • Body mass index • Socioeconomic status • Maternal diagnoses • Co-prescribing

		<p>some individuals from Medicaid population</p> <p>Sample size ~32,000 deliveries per year. ~ 75% of infants remain in the health plan</p>			<p>office</p> <p>Opportunity for medical record review - Yes</p>	<ul style="list-style-type: none"> • Folic acid • Over-the-counter medicines • Reproductive history
<p>Régie de l'assurance maladie du Québec (RAMQ)^(53, 54)</p>	<p>Since 1980 – recipients of social welfare</p> <p>Since 1997 – workers and their families not covered under private drug insurance</p>	<p>Country Canada</p> <p>Population-based – No Drug information for only recipients of social welfare and those who do not have private healthcare</p> <p>Sample size ~20,000 pregnancies per year</p>	<p>Dispensed prescription data</p>	<ul style="list-style-type: none"> • Live births • Stillbirths • Spontaneous losses • Elective terminations 	<p>Diagnoses recorded in the administrative databases of RAMQ and MED-ECHO</p> <p>Opportunity for medical record review - No</p>	<ul style="list-style-type: none"> • Smoking status • Alcohol consumption • Body mass index • Socioeconomic status • Maternal diagnoses • Co-prescribing • Folic acid • Over-the-counter medicines • Reproductive history
<p>Clalit Data Warehouse^(55, 56)</p>	<p>Since 1998</p>	<p>Country Israel</p> <p>Population based - No Members of the Southern district of Clalit Health Services - ~70% of women 15-49 years</p> <p>Sample size ~9,500 births per year</p>	<p>Dispensed prescription data</p>	<ul style="list-style-type: none"> • Live births • Stillbirths • Spontaneous losses • Elective terminations 	<p>Medical diagnoses during hospitalisation drawn directly from hospital records</p> <p>Opportunity for medical record review - Yes</p>	<ul style="list-style-type: none"> • Smoking status • Alcohol consumption • Body mass index • Socioeconomic status • Maternal diagnoses • Co-prescribing • Folic acid - some • Over-the-counter medicines • Reproductive history

Purpose built surveillance systems						
Slone Epidemiology Unit Birth Defects Study ^(23, 57)	Since 1976	Country United States and previously Canada Population based - Yes Sample size To date >40,000 women have been interviewed	Self-reporting via maternal telephone questionnaire (face to face interview up until 1998)	<ul style="list-style-type: none"> • Live births • Stillbirths • Spontaneous losses • Elective terminations 	Recorded by a paediatrician Opportunity for medical record review - Yes, with mothers permission	<ul style="list-style-type: none"> • Smoking status • Alcohol consumption • Body mass index • Socioeconomic status • Maternal diagnoses • Co-prescribing • Folic acid • Over-the-counter medicines • Reproductive history Opportunity to add additional interview questions relevant to a particular study.
National Birth Defects Prevention Study ^(58, 59)	Since 1997	Country United States Population based - Yes Sample size ~10% of annual US birth cohort	Self reporting of exposure by maternal assisted telephone interview between 6 weeks and 2 years after the expected date of delivery	<ul style="list-style-type: none"> • Live births • Stillbirths • Spontaneous losses • Elective terminations The capture of stillbirths and elective terminations varies by state Controls are live births only	Medical record extraction Collect data on a sample of specific birth defect types but not all Opportunity for medical record review - Yes	<ul style="list-style-type: none"> • Smoking status • Alcohol consumption • Body mass index • Socioeconomic status • Maternal diagnoses • Co-prescribing • Folic acid • Over-the-counter medicines • Reproductive history Opportunity to add additional interview questions relevant to a particular study
The Latin-American Collaborative Study of Congenital	Since 1967	Country 9 countries in South America	Self reported by the mother and collected by a trained	<ul style="list-style-type: none"> • Live births • Stillbirths • Spontaneous losses 	Identified from registered malformations	<ul style="list-style-type: none"> • Smoking status • Alcohol consumption • Body mass index

Malformations (ECLAMC) ^(60, 61)		<p>Population based - Yes</p> <p>Sample size ~150 - 200,000 births per year</p>	paediatrician during the puerperium	<ul style="list-style-type: none"> • Elective terminations 	<p>diagnosed at birth</p> <p>Opportunity for medical record review - No</p>	<ul style="list-style-type: none"> • Socioeconomic status • Maternal diagnoses • Co-prescribing • Folic acid • Over-the-counter medicines • Reproductive history <p>Collects data on 50 possible risk factors</p>
Spanish Collaborative Study of Congenital Malformations (ECEMC) ^(62, 63)	Since 1976	<p>Country Spain</p> <p>Population based - Yes</p> <p>Sample size ~87,000 births per year ~1,100-1,300 case-control pairs per year</p>	Maternal interviews with paediatricians within the first 3 days following delivery.	<ul style="list-style-type: none"> • Live births • Stillbirths • Spontaneous losses • Elective terminations 	<p>Diagnosed by paediatricians within the first 3 days of life.</p> <p>Opportunity for medical record review - No</p>	<ul style="list-style-type: none"> • Smoking status • Alcohol consumption • Body mass index • Socioeconomic status • Maternal diagnoses • Co-prescribing • Folic acid • Over-the-counter medicines • Reproductive history <p>>300 data points of information collected.</p>
Hungarian Case-control of Congenital Abnormalities Study ⁽⁶⁴⁾	<p>1980-1996</p> <p>Appears to be no longer recruiting - Emailed Professor Czeizel but did not get a response</p>	<p>Country Hungary</p> <p>Population based - Yes</p> <p>Sample size In 1996 ~22,843 cases and 38,151 controls</p>	Review of antenatal log book and medical records recorded by obstetrician, additional data requested by maternal questionnaire	<ul style="list-style-type: none"> • Live births • Stillbirths • Spontaneous losses • Elective termination - following a prenatal malformation diagnosis 	<p>Cases reported by a physician or paediatrician during first 3 months after birth or termination.</p> <p>Opportunity for medical record review - Yes, discharge summaries</p>	<ul style="list-style-type: none"> • Smoking status • Alcohol consumption • Body mass index • Socioeconomic status - employment status only • Maternal diagnoses • Co-prescribing • Folic acid • Over-the-counter medicines • Reproductive history

European Concerted Action on Congenital Anomalies and Twins (EUROCAT) ⁽⁶⁵⁻⁶⁷⁾	Since 1979	Country 20 European countries Population based - Yes Sample size ~1.7 million births per year	Varies by register – hospital records, GP records, pharmacy records, maternal interview Not all registers capture drug exposure data	<ul style="list-style-type: none"> • Live births • Stillbirths • Spontaneous losses • Elective terminations 	Largely reported by a physician to a local or national congenital anomaly register Opportunity for medical record review - varies by registry	<ul style="list-style-type: none"> • Smoking status • Alcohol consumption • Body mass index • Socioeconomic status • Co-morbidities • Co-prescribing • Folic acid • Over-the-counter medicines • Reproductive history All vary by register
<u>New data sources undergoing evaluation</u>						
Secure Anonymised Information Linkage Databank (SAIL) ⁽⁶⁸⁾	General practice data since 1992 Hospital admissions from 2004	Country Wales Population based – Yes Sample size ~44,000 pregnancies per year	Prescriptions issued by a GP	<ul style="list-style-type: none"> • Live births • Stillbirths • Spontaneous losses • Elective terminations 	Diagnoses recorded by a GP or paediatrician Opportunity for medical record review - Yes	<ul style="list-style-type: none"> • Smoking status • Alcohol consumption • Body mass index • Socioeconomic status • Maternal diagnoses • Co-prescribing • Folic acid – high dose only • Over-the-counter medicines • Reproductive history
German Pharmacoepidemiological Research Database ⁽⁶⁹⁾	Assessment of a mother-baby link in 2010	Country Germany Population based – No German statutory health insurances Sample size ~83,000 live births per year	Dispensation data of reimbursed drugs	<ul style="list-style-type: none"> • Live births • Stillbirths • Spontaneous losses • Elective terminations <p>Work is ongoing into the capture of pregnancy losses</p>	Under investigation but have access to hospital data and ambulatory physician visits Opportunity for medical record review - No	<ul style="list-style-type: none"> • Smoking status • Alcohol consumption • Body mass index • Socioeconomic status • Maternal diagnoses • Co-prescribing • Folic acid • Over-the-counter medicines • Reproductive history

Evaluation chez la Femme des Medicaments et de leurs RISque (EFEMERIS Database)⁽⁷⁰⁾	Since 2004	<p>Country France</p> <p>Population based - No Pregnant women in the Haute-Garonne department registered under general state coverage (~80% of the population)</p> <p>Sample size ~13,500 pregnancies per year</p>	Dispensed prescription data recorded to be sent to the French Health Insurance System Caisse Primaire d'Assurance Maladie (CPAM)	<ul style="list-style-type: none"> • Live births • Stillbirths • Spontaneous losses • Elective terminations 	<p>Recorded by physician during compulsory medical examinations at 8 days, 9 months and 3 years</p> <p>Prenatal diagnoses resulting in a termination are recorded by the antenatal diagnostic centre</p> <p>Opportunity for medical record review - No</p>	<ul style="list-style-type: none"> • Smoking status • Alcohol consumption • Body mass index • Socioeconomic status • Maternal diagnoses • Co-prescribing • Folic acid • Over-the-counter medicines • Reproductive history
---	------------	---	--	---	---	--

2.3 Discussion of alternative data sources

The review of the literature identified a large number of data sources being used for drug safety in pregnancy research. Based on the population captured and the type of data collected they can be grouped into three broad categories: population-based surveillance registers that rely on linked data sets, healthcare databases and purpose-built data sources such as case-control surveillance systems. Below, the key strengths and limitations of each type of data source are summarised.

Population-based surveillance registers

A key strength of population-based surveillance registers, such as those of the Nordic countries, is the mandatory reporting of all live- and stillbirths within a country or region. This results in the capture of exposure and outcome data from a representative sample of women and reduces concerns about the generalisability of study findings. One limitation, however, is that not all of these registers capture spontaneous pregnancy losses and induced terminations of pregnancy.

In the past, almost all data collected on first trimester drug exposure in these registers would have been based on maternal self-reporting during antenatal visits. Today, however, many have access to linked prescription data and the independent recording by the prescriber has the advantage of removing the possibility of recall bias. Capturing prescription data only does, however, mean over-the-counter exposures are not covered and there is a lack of information on whether the woman actually took the medicine and the precise timing of exposure. Studies using population based surveillance registers often identify congenital malformations from birth defect registers. As malformations are reported to these registers by physicians, midwives or paediatricians, the recording and reliability of the data is thought to be high.

Population-based surveillance registers have similar restrictions to pregnancy registries in terms of the volume of information that can be feasibly collected on covariates of interest, owing to the time available during an antenatal care

interview with a midwife. Whilst they all tend to collect data on maternal chronic diseases and co-prescribing, data on lifestyle factors such as alcohol intake, smoking status and body-mass-index is not always available.

Healthcare databases

Two main types of healthcare database were identified from the review of the literature; those that contain patient medical records and those that are based on administrative claims for reimbursement of medical treatment and prescriptions. Medical record databases such as the GPRD and The Health Improvement Network (THIN) capture data on a representative sample of the UK population in terms of age, sex and morbidity.⁽⁷¹⁾ The representative nature of the population captured by claims databases, however, varies by the type of insurance policy. The population of Kaiser Permanente, for example, has been found to be reasonably representative of the geographical areas that it covers, although the extremes of household income are thought to be under-represented. Tennessee Medicaid, however, is a US government-funded scheme and generally captures more mothers from populations with lower socio-economic status.⁽⁷²⁾

Electronic medical record data has the advantage of exposure information being recorded prospectively by the prescriber before the pregnancy outcome is known.⁽⁷¹⁾ Claims data from dispensing sources also has the added advantage that exposure classification is based on dispensed, rather than prescribed, prescriptions but however, uncertainty remains as to whether the medication was actually used.⁽⁷³⁾ Neither source captures information on over-the-counter exposures including standard dose (400µg) folic acid.

Identification of congenital malformations within healthcare databases is based on the presence of medical codes relating to either a diagnosis or treatment for a congenital malformation. The level of detail and completeness of the information available in these codes varies considerably. In primary care medical records, diagnoses made in a hospital setting will only be recorded in the database if the patient's GP chooses to enter the information received from a

specialist. Medical codes recorded for the purpose of administrative claims may lack detail and accuracy as they are recorded purely for the purpose of creating a bill for payment and therefore for the purposes of the database it is the procedure, rather than the diagnosis, that is of the greatest importance. Primary care medical record databases have the advantage of capturing all types of pregnancy outcome including spontaneous abortions and induced terminations of pregnancy which are not commonly available within administrative claims databases.

Within healthcare databases medical information is routinely recorded preventing the need for active follow-up as is required by pregnancy registries. Medical record databases have the benefit that an individual can only be lost to follow-up if they change GP practice or the GP practice stops contributing data to the database. This enables individuals to be followed for many years without any additional effort and makes it possible to identify malformations diagnosed later in life. Administrative claims databases, however, often have less follow-up time as individuals may change insurer when they move jobs or when they become pregnant which can reduce the availability of exposure and outcome data for research purposes.

Electronic medical records such as the GPRD contain information on smoking, alcohol and body mass index (BMI) although this information is not always complete and available for every patient.⁽⁷⁴⁾ Information on lifestyle factors is less likely to be recorded in claims databases,⁽⁷³⁾ owing to the purpose and nature of the database, although there are exceptions like Tennessee Medicaid, which contains data on smoking.

One recognised advantage of healthcare databases is the large number of individuals and pregnancies that they capture. Contrary to some belief, however, small sample sizes can still be a limitation and the ability to identify an association in these databases is dependent on the prevalence of the disease being studied and the frequency of prescribing.⁽⁹⁾

Data sources that capture a representative sample of the population, rather than only those with a particular disease or exposure enable the identification of multiple internal comparator groups that will have been recruited in the same way as those exposed to the product of interest.⁽⁹⁾ Depending on the exposure(s) of interest, these data sources may still be limited in terms of the number of individuals that are eligible for inclusion in any particular control group.

Case-control surveillance systems

Case-control surveillance systems are purpose-built data sources where cases and controls are recruited with the aim of the data being analysed using the case-control study design. The efficiency and statistical power resulting from the case-control study design are key strengths in enabling these data sources to be used to detect increases in risk for rare outcomes and malformation types.

One of the main limitations of case-control surveillance systems is the fact that exposure data is collected by maternal self reporting after the pregnancy outcome is known. This has the potential to introduce recall bias if there is differential reporting of exposure between women who had a pregnancy outcome with a congenital malformation and those who did not. In some circumstances attempts can be made to control for this by selecting malformed controls for the risk assessment studies; either those with chromosomal defects or those with a malformation other than the one(s) of interest and thought not to be associated with the exposure under study.

Systems that rely on maternal self-reporting do, however, have the advantage that they are able to collect data on all types of exposures including those issued in a hospital, bought over-the-counter or even borrowed from a friend or relative. A further strength is that there is the ability to extend or adapt the interview questionnaire to include questions on any potential confounding variables that may be associated with the particular exposures and outcomes of interest.⁽⁵⁷⁾

Case-control surveillance systems either recruit cases of congenital malformations directly from hospitals or birth defect registries where they have been reported and diagnosed by a paediatrician and often have the benefit of access to patient medical records with the mother's consent. Although some systems do capture stillbirths and induced terminations of pregnancy⁽⁶⁶⁾ no system captures spontaneous pregnancy losses.

Purpose-built case-control surveillance systems have a number of strengths for drug safety in pregnancy research but unfortunately they are expensive and often trade-offs have to be made in terms of the amount and level of detail of information collected and the time and cost required for data collection. There is also a need to limit the amount of information requested to minimise the burden on participants in order to maximise recruitment.

Other data sources

In addition to the data sources with systematic data collection outlined in Table 2.2, the review of the literature identified a number of publications by Teratology Information Services (TIS).⁽⁷⁵⁻⁷⁹⁾ The TIS recruit women who have voluntarily contacted them in search of information on the safety of a medicine they have used during pregnancy. Women who consent participate in a short telephone interview and are given a diary to record any further exposures. They are then contacted shortly after the expected date of delivery to obtain information on the pregnancy outcome. The voluntary nature of enrolment of women in these studies means they are subject to potential selection and self-referral biases and often the number of exposures captured for a particular product is small. The TIS are, however, valuable signal generating tools and they have the strength that information on a large number of potential confounding variables can be collected.

The International Clearinghouse for Birth Defects Surveillance and Research (ICBDSR) was one source that was identified that did not fit into a single data source category.^(80,81) The ICBDSR is affiliated with the World Health Organisation and aims to bring together a range of data sources being used for birth defect

research including congenital anomaly registries, case-control surveillance systems and national birth registers. A number of the data sources listed in Table 2.2 also contribute data and are members of the ICBDSR.⁽⁸²⁾

Conclusion

There is a large number of data sources, in addition to pregnancy exposure registries, being used to monitor the safety of medicine use during pregnancy. A number of data sources were identified that are currently undergoing review to determine their suitability to be used in this kind of research.^(68, 69, 70)

When thinking about the extent of evidence available from different studies it is important to remember that not all data sources will be capable of capturing all exposures. Partly this will be because some sources do not capture exposures in hospitals or over-the-counter medicine use but it will also result from differences in prescribing practices and the availability of products in different countries.⁽¹¹⁾ It is because of these geographical variations that relatively small surveillance systems⁽⁷⁰⁾ can be incredibly valuable as a means of monitoring *in utero* drug exposure and its effects.

Few data sources were identified that monitor exposure and pregnancy outcomes in less developed countries. The patient characteristics and medicines available to pregnant women in these countries are likely to differ considerably from other geographical areas and the findings from studies in more developed countries may therefore not be generalisable. In recent years attempts have been made to develop a pregnancy exposure registry evaluating the safety of anti-malarial drugs in malaria-endemic countries⁽⁸³⁾ but it is likely to be a long time before the healthcare systems of many of these countries have an automated system that can be utilised for drug safety in pregnancy research.

The fact the data sources identified as part of this review differed in terms of their strengths and limitations, highlights the benefit that a combined approach using a range of data sources could have in enhancing the extent of information available to women and healthcare professionals. It is important, however, that

only data sources that contain mostly reliable and accurate information are used and it is therefore crucial that a thorough evaluation is carried out of all the different aspects involved. This thesis will take one data source, the General Practice Research Database, and will evaluate its potential to be used as a tool for monitoring the safety of medicines used during pregnancy. The following chapter outlines the aims and objectives that will be addressed as part of this evaluation.

2.4 References

1. Mitchell AA. Systematic identification of drugs that cause birth defects - a new opportunity. *N Engl J Med.* 2003;349(26):2556-9.
2. U.S. Department for Health Human Services, Food Drug Administration, Center for Drug Evaluation Research, Center for Biologics Evaluation Research. Guidance for industry: Establishing pregnancy exposure registries. 2002. Available from: <http://www.fda.gov/downloads/ScienceResearch/SpecialTopics/WomensHealthResearch/UCM133332.pdf> [Accessed 12.03.2012].
3. Committee for Medicinal Products for Human Use (CHMP). Guideline on the exposure to medicinal products during pregnancy: need for post-authorisation data. (2005). Available from: http://www.ema.europa.eu/docs/en_GB/document_library/Regulatory_and_procedural_guideline/2009/11/WC500011303.pdf [Accessed on 10.01.2012].
4. Moore KL. *The Developing Human*. 4th ed: W.B. Saunders Philadelphia; 1988.
5. Cunningham M, Messenheimer J. Pregnancy Registries: strengths, weaknesses, and bias interpretation of pregnancy registry data. *Int Rev Neurobiol.* 2008;83:283-304.
6. Kennedy DL, Uhl K, Kweder SL. Pregnancy exposure registries. *Drug Saf.* 2004;27(4):215-28.
7. Covington DL, Tilson H, Elder J, et al. Assessing teratogenicity of antiretroviral drugs: monitoring and analysis plan of the Antiretroviral Pregnancy Registry. *Pharmacoepidemiol Drug Saf.* 2004;13(8):537-45.
8. Committee for Medicinal Products for Human Use (CHMP) Guideline on risk assessment of medicinal products on human reproduction and lactation: from data to labelling. 2008 Available from: http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003307.pdf [Accessed on 23.01.2012.]
9. Charlton RA, Cunningham MC, de Vries CS, et al. Data resources for investigating drug exposure during pregnancy and associated outcomes: the General Practice Research Database (GPRD) as an alternative to pregnancy registries. *Drug Saf.* 2008;31(1):39-51.
10. White AD, Andrews EB. The pregnancy registry program at Glaxo Wellcome Company. *J Allergy Clin Immunol.* 1999;103:S362-S3.
11. Wyszynski DF. Pregnancy exposure registries: Academic opportunities and industry responsibility. *Birth Defects Res A Clin Mol Teratol.* 2009;85:93-101.

12. Morrow J, Russell A, Guthrie E, et al. Malformation risks of antiepileptic drugs in pregnancy: a prospective study from the UK Epilepsy and Pregnancy Register. *J Neurol Neurosurg Psychiatry*. 2006;77(2):193-8.
13. Kendle International Inc. (2008) Bupropion pregnancy registry. Final Report 1 September 1997 through 31 March 2008. Available from: http://pregnancyregistry.gsk.com/documents/bup_report_final_2008.pdf [Accessed on 12.03.2012].
14. Roberts SS, Mallard AE, Ferber SM, et al. Does monetary compensation decrease losses to follow-up in pregnancy exposure registries? *Pharmacoepidemiol Drug Saf*. 2007;16(S1):S151.
15. Holmes LB, Baldwin EJ, Smith CR, et al. Increased frequency of isolated cleft palate in infants exposed to lamotrigine during pregnancy. *Neurology*. 2008;70(22):2152-8.
16. Tomson T, Battino D, French J, et al. Antiepileptic drug exposure and major congenital malformations: The role of pregnancy registries. *Epilepsy Behav*. 2007;11(3):277-82.
17. Tomson T, Battino D, Craig J, et al. Pregnancy registries: differences, similarities, and possible harmonization. *Epilepsia*. 2010;51(5):909-15.
18. Hernandez-Diaz S, Mittendorf R, Holmes LB. Comparative safety of topiramate during pregnancy. *Birth Defects Res A Clin Mol Teratol*. 2010;88(5):408.
19. Charlton RA, Cunnington M, Weil JG, et al. Regarding the need, or lack thereof, of a comparator group for pregnancy registries. *Birth Defects Res A Clin Mol Teratol*. 2009;85:808.
20. Cunnington MC, Weil JG, Messenheimer JA, et al. Final results from 18 years of the International Lamotrigine Pregnancy Registry. *Neurology*. 2011;76(21):1817-23.
21. Cole JA, Ephross SA, Cosmatos IS, et al. Paroxetine in the first trimester and the prevalence of congenital malformations. *Pharmacoepidemiol Drug Saf*. 2007;16:1075-85.
22. Kallen BA, Otterblad Olausson P. Maternal use of selective serotonin re-uptake inhibitors in early pregnancy and infant congenital malformations. *Birth Defects Res A Clin Mol Teratol*. 2007;79(4):301-8.
23. Louik C, Lin AE, Werler MM, et al. First-trimester use of selective serotonin-reuptake inhibitors and the risk of birth defects. *N Engl J Med*. 2007;356(26):2675-83.
24. Wurst KE, Poole C, Ephross SA, et al. First trimester paroxetine use and the prevalence of congenital, specifically cardiac, defects: a meta-analysis of epidemiological studies. *Birth Defects Res A Clin Mol Teratol*. 2010;88(3):159-70.
25. Einarson A, Pistelli A, DeSantis M, et al. Evaluation of the risk of congenital cardiovascular defects associated with use of paroxetine during pregnancy. *Am J Psychiatry*. 2008;165(6):749-52.
26. Bakker MK, Kerstjens-Frederikse WS, Buys CHCM, et al. First-trimester use of paroxetine and congenital heart defects: A population-based case-control study. *Birth Defects Res A Clin Mol Teratol*. 2010;88(2):94-100.
27. Howard TB, Tassinari MS, Feibus KB, et al. Monitoring for teratogenic signals: Pregnancy registries and surveillance methods. *Am J Med Genet C Semin Med Genet*. 2011;157(3):209-14.
28. Reis M, Kallen B. Delivery outcome after maternal use of antidepressant drugs in pregnancy: an update using Swedish data. *Psychol Med*. 2010;40(10):1723-33.
29. Nezvalova-Henriksen K, Spigset O, Nordeng H. Triptan exposure during pregnancy and the risk of major congenital malformations and adverse pregnancy outcomes: results from the Norwegian Mother and Child Cohort Study. *Headache*. 2010;50(4):563-75.

30. Irgens LM. The Medical Birth Registry of Norway. Epidemiological research and surveillance throughout 30 years. *Acta Obstet Gynecol Scand*. 2000;79(6):435-9.
31. Artama M, Gissler M, Malm H, et al. Nationwide register-based surveillance system on drugs and pregnancy in Finland 1996–2006. *Pharmacoepidemiol Drug Saf*. 2011;20(7):729-38.
32. Artama M, Auvinen A, Raudaskoski T, et al. Antiepileptic drug use of women with epilepsy and congenital malformations in offspring. *Neurology*. 2005;64(11):1874-8.
33. Molgaard-Nielsen D, Hviid A. Newer-generation antiepileptic drugs and the risk of major birth defects. *JAMA*. 2011;305(19):1996-2002.
34. Wallach Kildemoes H, Toft Sørensen H, Hallas J. The Danish National Prescription Registry. *Scand J Public Health*. 2011;39(7):38-41.
35. Christensen K, Knudsen L. Registration of congenital malformations in Denmark. *Dan Med Bull*. 1998;45:91-4.
36. Sorensen HT, Skriver MV, Pedersen L, et al. Risk of infantile hypertrophic pyloric stenosis after maternal postnatal use of macrolides. *Scand J Infect Dis*. 2003;35(2):104-6.
37. Wen SW, Yang Q, Garner P, et al. Selective serotonin reuptake inhibitors and adverse pregnancy outcomes. *Am J Obstet Gynecol*. 2006;194(4):961-6.
38. Downey W, Stang M, Beck P, et al. Health Services databases in Saskatchewan. In: BL S, editor. *Pharmacoepidemiology*. 4th ed: John Wiley & Sons, Ltd; 2005. p. 295-310.
39. Wang LH, Lin HC, Lin CC, et al. Increased risk of adverse pregnancy outcomes in women receiving zolpidem during pregnancy. *Clin Pharmacol Ther*. 2010;88(3):369-74.
40. Colvin L, Slack-Smith L, Stanley FJ, et al. Pharmacovigilance in pregnancy using population-based linked datasets. *Pharmacoepidemiol Drug Saf*. 2009;18:211-25.
41. Colvin L, Slack-Smith L, Stanley FJ, et al. Linking a pharmaceutical claims database with a birth defects registry to investigate birth defect rates of suspected teratogens. *Pharmacoepidemiol Drug Saf*. 2010;19(11):1137-50.
42. Gagne JJ, Maio V, Berghella V, et al. Prescription drug use during pregnancy: a population-based study in Regione Emilia-Romagna, Italy. *Eur J Clin Pharmacol*. 2008;64(11):1125-32.
43. Charlton RA, Weil JG, Cunningham MC, et al. Identifying major congenital malformations in the UK General Practice Research Database (GPRD): a study reporting on the sensitivity and added value of photocopied medical records and free text in the GPRD. *Drug Saf*. 2010;33(9):741-50.
44. Charlton RA, Weil JG, Cunningham M, et al. Comparing the General Practice Research Database and the UK Epilepsy and Pregnancy Register as tools for postmarketing teratogen surveillance: anticonvulsants and the risk of major congenital malformations. *Drug Saf*. 2011;34(2):157-71.
45. Bourke A, Dattani H, Robinson D. Feasibility study and methodology to create a quality-evaluated database of primary care data. *Inform Prim Care*. 2004;12(3):171-7.
46. Petersen I, Gilbert R, Evans J, et al. Pregnancy as a major determinant for discontinuation of antidepressants: An analysis of data from The Health Improvement Network. *J Clin Psychiatry*. 2011;72(7):979-85.
47. Cooper WO, Hernandez-Diaz S, Arbogast PG, et al. Antibiotics potentially used in response to bioterrorism and the risk of major congenital malformations. *Paediatr Perinat Epidemiol*. 2009;23(1):18-28.

48. Newschaffer CJ, Cocroft J, Anderson CE, et al. Prenatal zidovudine use and congenital anomalies in a medicaid population. *J Acquir Immune Defic Syndr*. 2000;24(3):249-56.
49. Li D-K, Liu L, Odouli R. Exposure to non-steroidal anti-inflammatory drugs during pregnancy and risk of miscarriage: population based cohort study. *BMJ*. 2003;327(7411):368.
50. Selby J, Smith D, Johnson E, et al. Kaiser Permanente Medical Care Program. In: Strom BL, editor. *Pharmacoepidemiology*. 4th ed: John Wiley & Sons Ltd; 2005. p. 241-60
51. Drinkard CR, Shatin D, Clouse J. Postmarketing surveillance of medications and pregnancy outcomes: Clarithromycin and birth malformations. *Pharmacoepidemiol Drug Saf*. 2000;9:549-56.
52. Shatin D, Rawson N, Stergachis A. UnitedHealth Group. In: Strom BL, editor. *Pharmacoepidemiology*. 4th ed: John Wiley & Sons Ltd; 2005. p. 271-80.
53. Blais L, Beauchesne MF, Lemiere C, et al. High doses of inhaled corticosteroids during the first trimester of pregnancy and congenital malformations. *J Allergy Clin Immunol*. 2009;124(6):1229-34 e4.
54. Berard A, Lacasse A. Validity of perinatal pharmacoepidemiologic studies using data from the RAMQ administrative database. *Can J Clin Pharmacol*. 2009;16(2):e360-9.
55. Matok I, Gorodischer R, Koren G, et al. The safety of H(2)-blockers use during pregnancy. *J Clin Pharmacol*. 2010;50(1):81-7.
56. Matok I, Gorodischer R, Koren G, et al. Exposure to folic acid antagonists during the first trimester of pregnancy and the risk of major malformations. *Br J Clin Pharmacol*. 2009;68(6):956-62.
57. Mitchell AA. Adverse drug reactions in utero: Perspectives on teratogens and strategies for the future. *Clin Pharmacol Ther*. 2011;89(6):781-3.
58. Yoon P, Rasmussen S, Lynberg M, et al. The National Birth Defects Prevention Study. *Public Health Rep*. 2001;116(S1):32-40.
59. Gilboa SM, Strickland MJ, Olshan AF, et al. Use of antihistamine medications during early pregnancy and isolated major malformations. *Birth Defects Res A Clin Mol Teratol*. 2009;85(2):137-50.
60. Castilla E, Orioli I. ECLAMC: The Latin-American Collaborative Study of Congenital Malformations. *Community Genet* 2004;7:76-94.
61. Rittler M, López-Camelo J, Castilla EE. Monitoring congenital rubella embryopathy. *Birth Defects Res A Clin Mol Teratol*. 2004;70(12):939-43.
62. Martínez-Frías ML. Postmarketing analysis of medicines: methodology and value of the spanish case-control study and surveillance system in preventing birth defects. *Drug Saf*. 2007;30(4):307-16.
63. Rodríguez-Pinilla E, Mejías C, Prieto-Merino D, et al. Risk of hypospadias in newborn infants exposed to valproic acid during the first trimester of pregnancy: A case-control study in Spain. *Drug Saf*. 2008;31(6):537-43.
64. Czeizel AE, Rockenbauer M, Siffel C, et al. Description and mission evaluation of the Hungarian case-control surveillance of congenital abnormalities, 1980-1996. *Teratology*. 2001;63(5):176-85.
65. Jentink J, Loane MA, Dolk H, et al. Valproic acid monotherapy in pregnancy and major congenital malformations. *N Engl J Med*. 2010;362(23):2185-93.
66. Dolk H. EUROCAT: 25 years of European surveillance of congenital anomalies. *Arch Dis Child Fetal Neonatal Ed*. 2005;90:F355-F8.
67. Greenlees R, Neville A, Addor MC, et al. Paper 6: EUROCAT member registries: Organization and activities. *Birth Defects Res A Clin Mol Teratol*. 2011;91(S1):S51-S100.

68. Lyons R, Jones K, John G, et al. The SAIL databank: linking multiple health and social care datasets. *BMC Med Inform Decis Mak.* 2009;9(1):3.
69. Garbe E, Suling M, Kloss S, et al. Linkage of mother–baby pairs in the German Pharmacoepidemiological Research Database. *Pharmacoepidemiol Drug Saf.* 2011;20(3):258-64.
70. Lacroix I, Hurault C, Sarramon MF, et al. Prescription of drugs during pregnancy: a study using EFEMERIS, the new French database. *Eur J Clin Pharmacol.* 2009;65(8):839-46.
71. Walley T, Mantgani A. The UK General Practice Research Database. *Lancet.* 1997;350(9084):1097-9.
72. Hennessy S, Carson J, Ray W, et al. Medicaid Databases. In: Strom B, editor. *Pharmacoepidemiology*. 4th ed: John Wiley & Sons Ltd; 2005.
73. Hennessy S. Use of healthcare databases in pharmacoepidemiology. *Basic Clin Pharmacol Toxicol.* 2006;98:311-3.
74. Ray WA. Improving Automated Database Studies. *Epidemiology.* 2011;22(3):302-4
75. Chambers CD, Braddock SR, Briggs GG, et al. Postmarketing surveillance for human teratogenicity: A model approach. *Teratology.* 2001;64:252-61.
76. Leen-Mitchell M, Martinez L, Gallegos S, et al. Mini-review: History of organized teratology information services in North America. *Teratology.* 2000;61(4):314-7.
77. Schaefer C, Hannemann D, Meister R. Post-marketing surveillance system for drugs in pregnancy-15 years experience of ENTIS. *Reprod Toxicol.* 2005;20(3):331-43.
78. Chambers CD, Johnson DL, Robinson LK, et al. Birth outcomes in women who have taken leflunomide during pregnancy. *Arthritis Rheum.* 2010;62(5):1494-503.
79. Diav-Citrin O, Shechtman S, Schwartz V, et al. Pregnancy outcome after in utero exposure to colchicine. *Am J Obstet Gynecol.* 2010;203(2):144 e1-6.
80. Clementi M, Di Gianantonio E, Cassina M, et al. Treatment of hyperthyroidism in pregnancy and birth defects. *J Clin Endocrinol Metab.* 2010;95(11):E337-41.
81. Botto LD, Robert-Gnansia E, Siffel C, et al. Fostering international collaboration in birth defects research and prevention: A perspective from the International Clearinghouse for Birth Defects Surveillance and Research. *Am J Public Health.* 2006;96(5):774-80.
82. International Clearinghouse for Birth Defects Surveillance and Research. Available from: <http://www.icbdsr.org/page.asp?p=12919&l=1> [Accessed on 06.03.2012].
83. Dellicour S, ter Kuile FO, Stergachis A. Pregnancy exposure registries for assessing antimalarial drug safety in pregnancy in malaria-endemic countries. *PLoS Med.* 2008;5(9):e187.

Chapter 3

Aims and objectives

Aim

Pregnancy exposure registries are commonly used to monitor for major teratogenicity following the introduction of a new medicine on the market. Over the last fifteen years, however, there has been an expansion in the range and number of data sources that are being used for the post-marketing evaluation of the safety of medicines when used during pregnancy. In particular, this has involved an increase in the use of automated healthcare databases including the United Kingdom's General Practice Research Database (GPRD). Preliminary investigations of pregnancy-related research using the GPRD have demonstrated its potential to be used in this field of research. These investigations have, however, often evaluated different aspects of the database in isolation. This PhD aims to evaluate the utility of the United Kingdom's General Practice Research Database to act as an alternative or complement to pregnancy registries in a more complete fashion. Given the known teratogenic effects of the older anticonvulsant drugs, the large amount of comparison data from anticonvulsant pregnancy registries, the chronic nature of epilepsy and the fact the majority of treatment will be prescribed within primary care; anticonvulsants were chosen for use as a case study.

Objectives

1. The ability to identify pregnancies and pregnancy outcomes is an essential requirement of a data source that can be used for drug safety in pregnancy research. The first objective was to identify pregnancies in the GPRD within a cohort of women with epilepsy.
2. Given the narrow time windows of the different stages of organ and tissue development, accurate information on the gestational timing of exposure to the product of interest is required. The second objective was to identify, within the cohort of women with a diagnosis of epilepsy, those who were exposed to anticonvulsants during the first trimester of pregnancy.

3. To enable the risk of a pregnancy outcome with a congenital malformation following a particular drug exposure to be evaluated, it needs to be possible to identify the offspring live deliveries. For the pregnancies identified in objective one, the third objective was to link, for live deliveries, the mother's medical record to that of the child.
4. The prevalence of all major congenital malformations is the outcome most commonly evaluated by pregnancy exposure registries. The fourth objective was to identify, within the cohort of women, all pregnancy outcomes with a major congenital malformation.
5. It is important the clinical information available in relation to major congenital malformations is accurate, as outcome misclassification can have a large impact on the calculated risk estimates. The fifth objective was to verify the malformations identified using either full photocopied medical records or free text comments recorded by GPs.
6. To assess the reliability of the GPRD as a tool for evaluating drug safety in pregnancy, the sixth objective was to compare findings from GPRD data with those from a pregnancy registry. To do this, the prevalence of major congenital malformations in the cohort of women identified was calculated and compared with those reported by the UK Epilepsy and Pregnancy Register in 2006.
7. Pregnancy exposure registries aim to detect new teratogens. A number of studies have concluded that the anticonvulsant valproate is a teratogen, with exposure during pregnancy increasing the risk of a pregnancy outcome with spina bifida. The seventh objective of this PhD was to determine whether the GPRD can be used to identify a known teratogenic association. To this end I focused on first trimester exposure to valproate in monotherapy and an increased risk of a pregnancy outcome with spina bifida.

8. The inclusion of pregnancies that do not result in a delivery is essential when evaluating the safety of a medicine when used during pregnancy. One of the advantages of the GPRD is that it captures the full range of pregnancy outcomes including spontaneous pregnancy losses and induced pregnancy terminations. The final objective was to aid the development of an algorithm that in addition to identifying pregnancy losses in the GPRD is also able to categorise them in an automated manner into those that were spontaneous, induced for a foetal medical condition and induced for other reasons.

Chapter 4

The General Practice Research Database

4.1 An overview of the General Practice Research Database

History

The General Practice Research Database (GPRD) contains anonymised, longitudinal, medical records routinely collected within UK general practice.^(1, 2) It was originally created in June 1987 as a tool to enable general practitioners (GPs) to record information relevant for patient care with the added benefit that the data was being held in a central system and could be anonymised and used for public health research.⁽³⁾ After a number of changes in management and ownership, the responsibility and management of the database are now with the GPRD division of the Medicines and Healthcare products Regulatory Agency (MHRA).⁽¹⁾

Data collection

The GPRD consists largely of coded data entered onto a computer system by GPs as part of the clinical management of patients within general practice. Information on patient demographics such as age and sex is also available. When a patient visits their GP, the date and type of consultation is recorded using the computer software along with information relating to symptoms, clinical diagnoses, detailed prescription data and some results of clinical investigations and tests. In the UK the GP acts as the gatekeeper to services within the National Health Service and therefore in addition to GP consultations the GP may also record information relating to hospital or specialist referrals and admissions as well as outpatient and emergency visits.^(4, 5) Patient anonymisation occurs within the GP practice at the point at which the data is downloaded to be transferred to the MHRA and results in each patient being assigned a unique identification number within each practice. Each GP practice also has a unique GP practice number so that the identity of the practices remains confidential.

Medical records

Data is largely entered onto the database in the form of medical codes. In 1995, following the donation of the database to the UK Department for Health, the computer software used by participating GP practices changed from a DOS-based

to a Windows-based platform. This resulted in a change in the coding system used for recording medical symptoms and diagnoses from OXMIS (Oxford Medical Information System) codes to Read clinical terms. As of 2007 all OXMIS codes entered onto the DOS-based version of the software have been cross-mapped to Read codes.

In addition to coded data, GPs have the option of recording un-coded comments, such as more detailed descriptions of symptoms, diagnoses or treatments along with information provided to them via hospital letters and discharge summaries. Owing to the need for anonymisation this so called 'free text' is not readily available to researchers and needs to be requested from the database provider at a fee.⁽⁶⁾ A large number of studies have been carried out to assess the accuracy and completeness of recording in the GPRD for a variety of diagnoses and these have been reassuring as estimates of validity have been high.^(7, 8, 9) Given the wide range of diagnoses recorded in the GPRD there are still, however, many diagnoses where no verification assessment has been carried out.

Prescription records

Prescription data has been found to be reasonably complete in the GPRD and this is likely to result from the GP actually needing to use the computer system in order to generate a prescription.⁽⁴⁾ Prescription information that can be recorded within the database includes the date the prescription was written, generic name and formulation, strength, quantity, daily dose and the duration of the prescription. Although the GPRD has the advantage that prescription information is recorded independently by the prescriber, prescription records in the GPRD refer only to prescriptions issued and not those that are necessarily dispensed or consumed. Over-the-counter medicines are not captured, nor are those that are issued during a hospital stay or by a consultant or specialist. Any outpatient prescriptions and refills/repeat prescriptions for those initiated in secondary care, however, are likely to be handled by the GP and should be captured.

Data quality

The recording of data from each GP practice is subject to quality control checks and each practice is assigned an 'up-to-standard' (UTS) date, which is the date the database provider considered the practice to have started contributing data that is of a standard suitable for the purposes of research. At the time of this study, assessment of being a UTS practice was based on 10 measures including the percentage of 'acceptable' patients³; the death, referral and monthly prescription rates; the percentage of prescriptions for which a medical indication has been recorded and the percentage of referrals with a clinical speciality.⁽¹⁾ Each patient is also assigned individual left and right censor dates, between which the data within their medical record is considered to be UTS. The left censor date (LCENS) is the later of the date the patient joined the practice or the date the practice was considered to be providing UTS data. The right censor date (RCENS) is the earliest of the date the patient left the practice, the date the practice stopped contributing data or was no longer considered to be contributing data that was UTS, the date the patient died or the date of the most recent data collection.

In general, only data recorded during UTS time periods is considered as eligible for inclusion in studies. When a patient registers at a practice for the first time, however, GPs may record details of historical medical events they consider relevant to the patient's future medical care, such as the date they were diagnosed with a particular medical condition (e.g. epilepsy or diabetes). These events will be recorded as occurring before a patient's LCENS. As GPs are not required to repeatedly record diagnoses this may be the only entry of a diagnosis within their medical record. Ignoring these records would restrict the study population to only those individuals who were diagnosed with a condition whilst registered at their current GP practice or had the diagnosis re-entered, so depending on the aim of the study and whether there is a requirement for

³ Those who are permanently registered with a valid first registration date and year of birth, no information recorded before the year of birth, a complete and valid transferred out date and reason where applicable, ≤ 115 years of age at end of follow-up and of known or indeterminate sex.⁽²⁾

incident rather than prevalent cases, there are situations where these records may need to be taken into account.

Links to other data sources

In recent years, attempts have been made by the database provider to link GPRD data with that of Hospital Episode Statistics (HES) and the National Cancer Registries. Both these links are still in the early stages and are restricted to those GP practices that consented to the further linkage (around 40%) and to data from 2005 onwards. In the latest GPRD dataset some aspects of the HES data for those practices consenting became freely available for the first time. Cancer registry data, however, still needs to be requested and these requests are assessed on a project by project basis.

Verification services

The nature of recording in the GPRD means that in some circumstances the presence of a single code for a particular condition may not be considered confirmation of a diagnosis. There may be scenarios when a GP suspects a particular diagnosis and records it in the patient's medical record but it is later diagnostically ruled out but not updated or only updated as a free text comment.⁽⁹⁾ As a result, the GPRD division at the MHRA offers a number of verification services to enable diagnoses identified via the computerised records to be confirmed or refuted. For all individuals, regardless of whether they are still registered with a GP who contributes data to the GPRD, it is possible to request additional anonymised information recorded by the GP in the free text fields as mentioned previously. If the patient is still registered with the practice a questionnaire can be sent to the GP, or less commonly to the patient themselves, to obtain additional information. For patients still registered it is also possible to request an anonymised photocopy of the patient's part or full medical record which enables access to all referral and outpatient letters and correspondence from consultants and specialists. All these services are available at a fee to the researcher and involve going through the GPRD division at the MHRA where the content is anonymised. If additional data is not requested it is possible to evaluate the reliability of the cases identified by comparing the rates

identified in the GPRD with those reported by other data sources. In addition the use of algorithms which require supporting evidence (e.g. surgery or the issue of a prescription) can also be used in an attempt to minimise case misclassification.^(7, 9)

Ethical approval

All studies using the GPRD that are carried out with the aim of publication, or communicating results to third parties, must receive International Scientific Advisory Committee (ISAC) approval before proceeding. The Multi-center Research Ethics Committee (MREC) has given approval for all purely observational GPRD studies. Investigators conducting studies that require any form of direct patient involvement do, however, need to apply for separate MREC approval.⁽²⁾

Viewing patients' medical records

The Pharmacoepidemiology unit at the University of Bath has developed software, referred to as 'the browser', which enables researchers to view/browse a patient's electronic medical record in chronological order. Reviewing patients' medical records in this way can prove very valuable for getting an overall impression of a patient's medical condition and the type of codes and data that are recorded. It is possible to review all clinical, test and therapy records together and also to review records masked to particular medical codes and therapy records to ensure that the reviewer makes this assessment regarding either exposure or outcome status independent of knowledge of the other. This tool has been very beneficial in creating and refining algorithms for identifying pregnancies in the GPRD as will become apparent throughout this thesis. An example of a patient's medical record as viewed in the browser is shown in Figure 4.1.

View Patient Medical History [F024]			
Patient ID	21916	Date of Birth	01/01/1971
Reg Date	10/09/1993	Lcens	10/09/1993
Transfer Date		Rcens	24/02/2006
		Practice	804038
		Death Date	
List			
Full History	Clinical Group	Therapy Group	
19/11/2002	FOLIC ACID tabs 400micrograms (02420001) [31]		
19/11/2002	Consultation: Other () [31]		
29/11/2002	Consultation: Repeat Issue () [31]		
15/01/2003	Last menstrual period -1st day (276157) [32]		
15/01/2003	Maternity: EDD:22-OCT-2003WkGest: Last menstrual period -1st day (276157) Births: Miscar: [32]		
20/02/2003	Consultation: Telephone call from a patient () [32]		
20/02/2003	Had a chat to patient (250993) [32]		
28/02/2003	EPANUTIN caps 50mg (01684002) [32]		
28/02/2003	FOLIC ACID tabs 5mg (02863001) [32]		
28/02/2003	Maternity: EDD:WkGest: Delivery booking place (204092) Births: Miscar: [32]		
28/02/2003	Patient pregnant (231247) [32]		
28/02/2003	EPANUTIN caps 100mg (01684003) [32]		
28/02/2003	Consultation: Surgery consultation () [32]		
28/02/2003	Delivery booking place (204092) [32]		
03/03/2003	Consultation: Other () [32]		
03/03/2003	O/E - blood pressure reading (285459) [32]		
03/03/2003	Referral: Hospital Obstetrics (Out Patient/Obstetrics) (231247) Urgency: Routine Patient pregnant [32]		
03/03/2003	Blood Pressure: Syst: 120 Dias: 80 (285459) [32]		
05/03/2003	O/E - blood pressure reading (285459) [32]		
05/03/2003	Antenatal examinations (258663) [32]		
05/03/2003	Consultation: Clinic () [32]		
05/03/2003	Obstetric examination (230627) [32]		
05/03/2003	Blood Pressure: Syst: 100 Dias: 60 (285459) [32]		
10/04/2003	Consultation: Letter from Outpatients () [32]		
22/04/2003	Consultation: Letter from Outpatients () [32]		
06/05/2003	FOLIC ACID tabs 5mg (02863001) [32]		
06/05/2003	EPANUTIN caps 50mg (01684002) [32]		
06/05/2003	EPANUTIN caps 100mg (01684003) [32]		
06/05/2003	Consultation: Repeat Issue () [32]		
07/05/2003	Consultation: Clinic () [32]		
07/05/2003	Antenatal examinations (258663) [32]		
07/05/2003	Test: Urinalysis - Glucose Urine glucose test negative (295019) Value: No Data Enter Range: (-) [32]		
07/05/2003	Obstetric examination (230627) [32]		

Figure 4.1 A pregnant patient's medical record as viewed in the browser

GPRD population

Since its creation in 1987 the number of patients contributing data to the GPRD has gradually grown. As of October 2011 the GPRD contained over 68 million person years of data and was actively collecting data on approximately 5 million patients (~8% of the UK population) registered at around 630 GP practices within the UK.⁽²⁾ The GPRD contains over 1.5 million pregnancies that ended between January 1992 and December 2010 and has the strength of capturing all types of

pregnancy outcome including live births, stillbirths, spontaneous abortions and induced terminations. Of the pregnancies recorded in the GPRD approximately 73% result in a delivery and 27% in a pregnancy loss.^(10, 11)

4.2 Features of the GPRD specific to drug safety in pregnancy research

When using a database to carry out drug safety in pregnancy research, the methods involved in the identification of eligible pregnancies, exposure classification and identification of outcomes, differ from those of pregnancy exposure registries and case-control surveillance systems. The exact methods are also likely to vary depending on the database being used. The next section summarises some of the features specific to drug safety in pregnancy research using the GPRD.

Identification of pregnancies

Identifying pregnancies in the GPRD requires the consideration of a large range of pregnancy related medical codes including those referring to pregnancy tests, the date of the first day of the last menstrual period, estimated date of delivery, delivery bookings, antenatal care, pregnancy outcomes including deliveries, stillbirths, spontaneous pregnancy losses and induced pregnancy terminations, as well as records of neonatal and postnatal care. For the pregnancy algorithm developed at the University of Bath, initially all pregnancy related medical codes are identified and each code is categorised according to whether it provides information on a delivery, a pregnancy loss, the gestational age at pregnancy outcome (e.g. premature delivery), a postpartum event or less specific supporting evidence of a pregnancy (e.g. an antenatal care record). Each record is then assessed in terms of the level of evidence it provides; so for example a record of a 'normal delivery' would be considered sufficient evidence, whilst a female with a record relating to the date of her last menstrual period would require additional medical codes providing supporting evidence of a pregnancy. All records of a negative pregnancy test are excluded from this process. For patients who have a medical record that requires supporting evidence, checks are made to ensure an additional code is present within a specified time frame.

Once this has been completed, all pregnancy outcome events are identified and categorised as being either a delivery (including stillbirth) or a pregnancy loss.⁽¹⁰⁾

Determining pregnancy end dates

It is possible that GPs may record duplicate pregnancy outcome records on different dates within the GPRD, for example one record on the date of delivery and another record when the female attends her six week post-natal visit. It is therefore necessary to estimate the actual pregnancy end date and this is important, as will be discussed in the next section, because in many cases it plays an essential role in determining the start date of the pregnancy and subsequent timings of exposure. The method used at the University of Bath to determine pregnancy end dates differs depending on whether the pregnancy resulted in a delivery or a pregnancy loss. For pregnancies that ended in a delivery, all pregnancy outcome records are listed in ascending date order and the earliest delivery event within each group of records for a patient is taken as the end date of that pregnancy. For pregnancy loss events, following listing them in ascending date order, the last event in a series of six weeks is taken as the pregnancy end date. This is because the earlier pregnancy loss codes commonly relate to requests and referrals for a termination rather than the date the pregnancy actually ended.

Determining pregnancy start dates

When the outcome of drug safety in pregnancy research is the risk of congenital malformations, accurately determining the pregnancy duration and timing of drug exposure is crucial given the narrow time window when key foetal development takes place and drug exposures have the opportunity to interfere and cause adverse effects.⁽¹²⁾ For deliveries, where recorded, information on the expected date of delivery or the date of the last menstrual period is used to calculate the start date of each pregnancy. Where these are not present, the algorithm identifies records relating to gestational age (e.g. antenatal care - 32 weeks) or the existence of a time-related code denoted as premature or post-mature (e.g. baby premature 36 weeks) and these are used to determine the start of the pregnancy. If none of these are recorded, the pregnancy is assigned a

default duration of 40 weeks. For pregnancies that end in a pregnancy loss, the start date of the pregnancy is determined, where possible, using the same method as for the deliveries, however a duration of 10 weeks is assigned where no information is available regarding gestational age.

There are some pregnancies in the GPRD where codes that are taken as indicative of a delivery (e.g. 'delivery details' and 'birth details') are recorded alongside or in close proximity to codes relating to a miscarriage or spontaneous loss. Manual review of a sample of these patients' medical records in the browser made it apparent there were instances where GPs recorded codes that could imply the delivery of a live born infant when actually they were recording details relating to a pregnancy loss. Where a record for a delivery has a code for a spontaneous or induced pregnancy loss within the 42 days before or 21 days after, the delivery record is disregarded.

Determining the type and timing of exposure

Prescription records within the GPRD relate to the issue of a prescription. Given the timing of drug exposure is so critical to the identification of possible teratogenic effects, it is not considered appropriate to establish exposure status based on prescriptions that were issued only during the actual pregnancy. This is because a prescription with a 60 day duration issued 15 days before the pregnancy start date, if taken in full, will result in the female being exposed during the first 45 days of her pregnancy, which is a time when central nervous system and cardiac development is already taking place. When determining first trimester drug exposure all prescriptions issued in the four to six months before pregnancy should also be identified and the exposure mapped based on their assumed duration. One limitation of using prescription records, however, is that women may decide not to take the medicines if they are trying to become pregnant, or once they discover they are pregnant, and using GPRD data it is not possible to identify or incorporate this discontinuation.

Linking the mother's record to the child's record

To be able to evaluate the safety of medicine use during pregnancy, in terms of the risk of congenital malformations in the offspring, it is necessary to link the mother's medical record to the record of the child. In addition to a patient identifier, individuals within the GPRD are also given a family number, which is primarily based on postal address. For each delivery, attempts are made to link the offspring to the mother based on a birth or registration record within the same family. Where two females with the same family number have pregnancy related codes recorded within two years of the birth, no attempts are made to link mother and child. Links were also not attempted where more than one possible child existed but with different dates of birth and where the mother might be in an institution (family size >20). Within the GPRD it is possible to link approximately 80% of all live pregnancy outcomes to a child.

Availability of information on potential risk factors and confounders

Within the GPRD it is possible to obtain information on a range of variables that are either risk factors for congenital malformations or could potentially act as confounders within drug safety in pregnancy research. These include maternal age, gestational age at pregnancy outcome, parity, gravidity, body mass index, alcohol status, smoking status, socioeconomic status, co-morbidities (e.g. diabetes) and co-prescribing. The completeness and accuracy of some of these variables have not, however, been officially validated. For some patients where information is missing it is possible to make assumptions, for example if someone was a non-smoker before they became pregnant it is considered unlikely that they will have taken up smoking during pregnancy. There are, however, a number of individuals who have no data recorded in their medical record relating to some or all of these variables and then their status is classified as unknown. Within the GPRD there will be some information on family history of congenital malformations, however this is thought to be selectively recorded and therefore not considered complete enough to provide reliable information. Data relating to genetics and over-the-counter prescribing is not available within the database.

4.3 Overview of drug safety in pregnancy research using the GPRD to date

The first studies that utilised the GPRD for drug safety in pregnancy research were published in the late 1990s. Since then a number of studies have been published, including papers reporting on methodological aspects of using the database in this way. The next section provides a summary of the key papers that have been published to date and aims to demonstrate where this PhD fits in terms of the broader picture of the GPRD as a tool for the post-marketing surveillance of potential new teratogens.

First use of the GPRD for drug safety in pregnancy research

The first studies to report on pregnancy outcome research using the GPRD were published by Jick *et al.* in the late 1990s.^(13, 14) These evaluated exposure to anticonvulsants⁽¹³⁾ and fluconazole,⁽¹⁴⁾ an antifungal agent, and the associated risk of having a child born with a congenital malformation. The methodologies reported for both these studies lacked detail on the exact methods of determining first trimester exposure and identification of congenital malformations. In addition both studies only captured live born pregnancy outcomes and had small sample sizes meaning they lacked statistical power. The study investigating anticonvulsants was, however, promising in that it found 100% agreement between data in the computerised medical records on the presence of a congenital malformation and the information received from GPs following the completion of a questionnaire. These studies were then followed by one investigating pregnancy outcomes after intra-uterine exposure to any of three acid suppression drugs.⁽¹⁵⁾ The methodology reported was more detailed than the papers by Jick *et al.* but, given what is known about the importance of the timing of drug exposure for this kind of research, the study had a wide time window of interest (30 days before through to 100 days after LMP). It did, however, capture non-live pregnancy outcomes including terminations of pregnancy.

Methodological studies

In 2004 the first study focussing on the methodological aspects of using the GPRD for drug safety in pregnancy research was published. This paper evaluated strategies for identifying pregnancies on the database with a focus on the complexity and the vast number of Read/OXMIS codes available for use by GPs.⁽¹⁶⁾ This area of research was then also covered by Snowball and de Vries in 2007⁽¹⁰⁾ who reported that between 1st January 1992 and 29th March 2006 they identified 494,449 pregnancies in the GPRD for women aged >11 years at the pregnancy outcome date, of which 72.8% resulted in a delivery and 27.2% resulted in a termination or miscarriage. In 2010 Devine *et al.* published a paper continuing on the topic of identifying pregnancies in the GPRD.⁽¹¹⁾ This paper not only reported on the creation of a computer based algorithm for identifying pregnancies but also on the results of validating the algorithm. The results of the validation demonstrated that the algorithm was more successful for live births but for other types of pregnancy, such as spontaneous abortions and induced terminations, further work was required as it is more difficult to determine pregnancy start and end dates when no child's record exists. Also often multiple records relating to a termination of pregnancy (TOP) exist.

Hardy *et al.* in 2006 went on to evaluate the types of drugs that women in the GPRD were prescribed during pregnancy and focussed on those pregnancies where it was possible to link the mother to her child.⁽¹⁷⁾ This study captured prescriptions received during two time periods; the 90 days before and the 70 days after the first recorded pregnancy code. Sixty-five percent of women received at least one prescription during one of the two time periods. Although more work is required to improve the precision of the timing of exposure, this study did highlight the large proportion of women within the GPRD who are exposed to drugs where the safety of the medication, when used during pregnancy, is not fully known.

Identification of congenital malformations

In 2007, the first research was published looking at the recording and verification of congenital malformations in the GPRD with a focus on cardiovascular

defects.^(18, 19) The first of these two studies compared the prevalence of congenital heart defects in the GPRD with two UK population based sources (the National Congenital Anomaly System and the UK contributors to EUROCAT)⁽¹⁹⁾ and the second study attempted to verify three specific heart defects identified in the electronic medical records, using additional information obtained from the free text and GP questionnaires.⁽¹⁸⁾

It was observed that the overall prevalence of congenital heart defects in the GPRD was higher than that reported in both of the UK population based sources. The authors suggested this could be owing to differences in data collection methods and in particular the voluntary nature of reporting by healthcare professionals to the UK population based sources. They concluded that the GPRD should capture the majority of cases for the population covered and has the advantage of including minor malformations that may not always be reported to the national registries. They also reported that the extended period of follow up in the GPRD is an advantage over sources such as EUROCAT, which although accept malformations diagnosed at any age, tend to focus on those diagnosed within the first year of life.⁽¹⁹⁾

In the verification study⁽¹⁸⁾ there was a 93.5% positive predictive value between the computerised records and responses to the GP questionnaires. This ranged from 90% for Tetralogy of Fallot to 100% for coarctation of the aorta. Fifty-one percent of infants were found to have free text comments in their medical records that provided additional information such as the findings from echocardiograms and past or future surgery. The authors concluded that the computerised medical records within the GPRD were suitable for studying the occurrence of these particular heart defects, but that where possible all infant free text attached to a malformation diagnosis should be requested to obtain as much detail as possible.

Similar research was published in 2008, this time focussing on neural tube defects (NTDs).⁽²⁰⁾ Owing to the nature of this class of defect this study included diagnoses of malformations that resulted in a termination of pregnancy, in

addition to those found in live and stillbirths. An algorithm was developed to identify new cases of NTDs in the GPRD in either the mother's or the child's record. The identified cases were then verified using GP questionnaires. The study identified 217 unique NTD cases; 148 recorded in the mother's record and 69 in the child's. Of the 169 questionnaires that were returned 117 (71%) NTD cases were verified. The positive predictive value (PPV) of the algorithm was found to vary by NTD diagnosis and ranged from 0.81 for anencephaly to 0.47 for spina bifida. The authors reported that the low PPV for spina bifida could have resulted from the inability of the algorithm to differentiate between cases where the mother herself had spina bifida and cases where it was present in the foetus or child. The annual prevalences of NTDs in the GPRD were comparable to those of the National Congenital Anomaly System in later years, but slightly higher in the early years. The authors concluded NTDs can be identified in the GPRD, therefore the GPRD provides an opportunity to evaluate maternal exposures and this severe class of congenital malformation. It was proposed that searching the free text comments or sending out a more comprehensive questionnaire may be required in order to minimise the false positives identified for cases of spina bifida.

Potential of the GPRD as a tool for post-marketing teratogen surveillance

In 2008 Charlton *et al.* published the paper referred to earlier which reported on the mean number of annual exposures to a range of drugs during pregnancy captured in the GPRD and made comparisons with the numbers enrolled in pregnancy exposure registries.⁽²¹⁾ This study found the GPRD has a potential role in the field of drug safety in pregnancy research and that for more prevalent conditions, such as depression, the GPRD should capture a similar mean annual number of exposures as a pregnancy registry. Since then, to our knowledge one manuscript has been published that reported on congenital malformations as an outcome following a relatively common exposure.⁽²²⁾ For drugs used to treat a disease with a relatively low prevalence and for those new on the market, however, pregnancy exposure registries were found to capture more exposures.

This brief review of the papers published to date has demonstrated that work is continuously ongoing in relation to using the GPRD for drug safety in pregnancy research. Over the last 10–15 years several attempts have been made to assess the reliability of the GPRD as a tool for post-marketing teratogen surveillance, in addition to trying to improve the quality and robustness of study findings. To date, however, the majority of work has been carried out to evaluate different aspects of the database in isolation and has focussed on either the methodological aspects of identifying pregnancies on the database or has been limited to the verification of a small number of specific congenital malformations. The work presented in the rest of thesis aims to evaluate all aspects of the GPRD relevant to drug safety in pregnancy research in combination. This will be done by conducting a risk assessment study and for the reasons outlined in the next chapter, uses anticonvulsants as a case study.

4.4 References

1. Wood L, Martinez C. The General Practice Research Database: Role in pharmacovigilance. *Drug Saf.* 2004;27(12):871-81.
2. Medicines and Healthcare products Regulatory Authority. The General Practice Research Database. (2011) Available from: <http://www.gprd.com/home/> [Accessed on 11.10.2011].
3. Wood L, Coulson R. Revitalizing the General Practice Research Database: plans, challenges, and opportunities. *Pharmacoepidemiol Drug Saf.* 2001;10:379-83.
4. Walley T, Mantgani A. The UK General Practice Research Database. *Lancet.* 1997;350(9084):1097-9.
5. Garcia Rodriguez LA, Perez-Gutthann S, Jick S. The UK General Practice Research Database. In: Strom BL, editor. *Pharmacoepidemiology*: John Wiley & Sons Ltd; 2000. p. 375-85.
6. Charlton RA, Weil JG, Cunningham M, et al. Identifying major congenital malformations on the General Practice Research Database: a study reporting on the sensitivity and added value of photocopied records and free text. *Drug Saf.* 2010;33(9):741-50.
7. Herrett E, Thomas SL, Schoonen WM, et al. Validation and validity of diagnoses in the General Practice Research Database: a systematic review. *Br J Clin Pharmacol.* 2010;69(1):4-14.
8. Khan NF, Harrison SE, Rose PW. Validity of diagnostic coding within the General Practice Research Database: a systematic review. *Br J Gen Pract.* 2010;60:e128-e36.

9. Charlton RA, Snowball JM, Bloomfield K, et al. Colorectal cancer incidence on the General Practice Research Database. *Pharmacoepidemiol Drug Saf.* 2012. *In Press*
10. Snowball JM, de Vries CS. Determination of pregnancy on the General Practice Research Database. *Pharmacoepidemiol Drug Saf.* 2007;16:S118.
11. Devine S, West S, Andrews E, et al. The identification of pregnancies within the General Practice Research Database. *Pharmacoepidemiol Drug Saf.* 2010;19(1):45-50.
12. Toh S, Mitchell AA, Werler MM, et al. Sensitivity and specificity of computerized algorithms to classify gestational periods in the absence of information on date of conception. *Am J Epidemiol.* 2008;167(6):633-40.
13. Jick S, Terris BZ. Anticonvulsants and congenital malformations. *Pharmacotherapy.* 1997;17(3):561-4.
14. Jick SS. Pregnancy outcomes after maternal exposure to Fluconazole. *Pharmacoepidemiol Drug Saf.* 1999;19(2):221-2.
15. Ruigomez A, Garcia Rodriguez LA, Cattaruzzi C, et al. Use of Cimetidine, Omeprazole and Ranitidine in pregnant women and pregnancy outcomes. *Am J Epidemiol.* 1999;150:476-81.
16. Hardy JR, Holford TR, Hall GC, et al. Strategies for identifying pregnancies in the automated medical records of the General Practice Research Database. *Pharmacoepidemiol Drug Saf.* 2004;13:749-59.
17. Hardy JR, Leaderer BP, Holford TR, et al. Safety of medications prescribed before and during early pregnancy in a cohort of 81 975 mothers from the UK General Practice Research Database. *Pharmacoepidemiol Drug Saf.* 2006;15:555-64.
18. Wurst KE, Ephross SA, Loehr J, et al. The utility of the General Practice Research Database to examine selected congenital heart defects: a validation study. *Pharmacoepidemiol Drug Saf.* 2007;16:867-77.
19. Wurst KE, Ephross SA, Loehr J, et al. Evaluation of the General Practice Research Database congenital heart defects prevalence: Comparison to United Kingdom national systems. *Birth Defects Res A Clin Mol Teratol.* 2007;79:309-16.
20. Devine S, West SL, Andrews E, et al. Validation of neural tube defects in the full featured - General Practice Research Database. *Pharmacoepidemiol Drug Saf.* 2008;17:434-44.
21. Charlton RA, Cunningham MC, de Vries CS, et al. Data resources for investigating drug exposure during pregnancy and associated outcomes: the General Practice Research Database (GPRD) as an alternative to pregnancy registries. *Drug Saf.* 2008;31(1):39-51.
22. Chi CC, Mayon-White RT, Wojnarowska FT. Safety of topical corticosteroids in pregnancy: A population-based cohort study. *J Invest Dermatol.* 2011;131(4):884-91.

Chapter 5

Anticonvulsants as a case study

Anticonvulsants are widely used for the treatment of epilepsy in addition to other conditions including bipolar disorder and neuropathic pain. As will be discussed below, there are several reasons why this class of drug has been extensively monitored by multiple pregnancy exposure registries and is considered a suitable case study for this PhD.

5.1 Treatment of epilepsy

Epilepsy is a medical condition that is defined as “a disorder of brain function characterised by recurrent seizures that have sudden onset”.⁽¹⁾ In approximately 75% of individuals, general control of epilepsy is achieved by taking anticonvulsant drugs.⁽²⁾ The broad spectrum of epileptic seizures and the fact they can occur anywhere in the brain, means there are a large number of different anticonvulsant products; each one often being more effective in treating some seizure types than others. In addition, not all individuals obtain sufficient seizure control from a single anticonvulsant and in approximately 20% of cases polytherapy treatment with a combination of anticonvulsant drugs is required. Given the likelihood of recurrent seizures and the potential harm these can cause to the individual, it is often inadvisable to stop anticonvulsant treatment. It is this inability to discontinue treatment that has a number of implications for women of child-bearing age.

5.2 Epilepsy and pregnancy

The incidence of epilepsy is highest in the first decade of life with a second peak occurring in those over 60 years of age.⁽³⁾ Although some individuals do not experience seizures as an adult and only suffer from childhood epilepsy, the young age at onset means that approximately a quarter of individuals with epilepsy are women of child bearing age.⁽⁴⁾ The number of pregnancies reported in the literature to be to women with epilepsy ranges from 3-6 per 1000⁽⁵⁻⁷⁾ depending on the country of study and whether stillbirths and pregnancy losses are included. In the UK, approximately 2,500 women with epilepsy will have a baby each year.⁽⁸⁾

There is some uncertainty over the role epilepsy as a condition plays in terms of being a risk factor for adverse pregnancy outcomes. As the number of women with epilepsy not prescribed anticonvulsants is small, and they may not choose to enrol in a registry, the effect of epilepsy itself is difficult to determine. A meta-analysis of ten studies that had captured a total of 400 women with epilepsy unexposed to anticonvulsants found no evidence to suggest epilepsy itself is associated with an increased risk of major congenital malformations when compared to non-epileptic controls.⁽⁹⁾ As women with untreated epilepsy are likely to have a less severe form of the disease and will also have better seizure control, questions remain over the comparability of women with treated and untreated epilepsy⁽¹⁰⁾ and given these differences, it is difficult to evaluate with any certainty the role of epilepsy severity and seizures on the risk of a pregnancy outcome with a major congenital malformation.

5.3 Anticonvulsant exposure and pregnancy outcomes

Post-marketing studies have shown that exposure during the first trimester of pregnancy, to the older generations of anticonvulsants, is associated with a two- to three-fold increased risk of a pregnancy outcome with a major congenital malformation, when compared to the risk in the general population.^(11, 12) The risk of congenital malformations has also been found to be higher in women exposed to polytherapy anticonvulsant treatment compared to those taking monotherapy.⁽¹¹⁾ In addition, studies have shown evidence of a dose response relationship associated with some anticonvulsants, with the risk of a congenital malformation increasing with increasing daily dose.^(11,13,14) In recent years, studies evaluating the effect of maternal anticonvulsant exposure on the neurodevelopment of the infant have found evidence that infants exposed *in-utero* to valproate or carbamazepine are more likely to have impaired cognitive function and developmental delay than those unexposed or taking other anticonvulsants such as lamotrigine.^(15, 16) As a result of all these factors, women with epilepsy who are considering becoming pregnant, are recommended to seek advice so attempts can be made to ensure they are on the optimum treatment regimen and exposed to the minimum number of anticonvulsants at the lowest effective dose.^(17, 18) Some enzyme inducing anticonvulsants, such as

carbamazepine and phenobarbital, however, have been found to interact with and reduce the effectiveness of hormonal contraceptives.⁽⁴⁾ This therefore has the potential to increase the likelihood of an unplanned pregnancy if alternative or additional methods of contraception are not used.

Even when pregnancies are planned, it is not always possible for women to switch to what is considered to be a safer product or lower dose. For example, a study by the EURAP Epilepsy and Pregnancy Register found the percentage of women who were seizure-free during pregnancy was lower in women taking lamotrigine monotherapy compared to valproate monotherapy.⁽¹⁹⁾ When making decisions about prescribing anticonvulsant therapy to women who are planning to become pregnant, it is important to balance the risks of potential adverse pregnancy outcomes with the benefits of having adequate seizure control. This is, however, complicated by the limited amount of information that is available on the teratogenic potential of some of the newer anticonvulsants owing to often small sample sizes in follow-up studies/pregnancy registries.

Of the older anticonvulsants, valproate and carbamazepine are the most commonly prescribed although both are associated with an increased risk of neural tube defects and in particular spina bifida. The risk of spina bifida following first trimester exposure to valproate is approximately 2%^(13,20) and following carbamazepine exposure is 1%⁽²¹⁾ compared to 0.1% in the general population. Studies in the general population and women who have previously had a pregnancy with a neural tube defect, have demonstrated that folic acid supplementation before and during the early stages of pregnancy is effective in reducing the risk of a pregnancy outcome with a neural tube defect.^(22,23) As a result all women planning to become pregnant, regardless of anticonvulsant exposure, are recommended to take 400mcg of folic acid daily before conception and until week 12 of pregnancy. Women considered to be at an increased risk, who have a history of a pregnancy or family member with a neural tube defect, are recommended to take a higher daily dose of 5mg. As some anticonvulsants including carbamazepine, phenytoin and phenobarbital are folic acid antagonists⁽²⁴⁾ women on anticonvulsant therapy are also recommended to take

the higher 5mg daily dose of folic acid before and during the first trimester of pregnancy.⁽²⁵⁾ The extrapolation of the protective and beneficial effect of folic acid supplementation in the general population to women taking anticonvulsants has, however, been questioned. A number of studies^(24, 26, 27) and case-reports⁽²⁸⁾ evaluating the effect of folic acid exposure early on in pregnancies exposed to anticonvulsant therapy have reported that they did not observe a reduction in the risk of neural tube defects and other congenital malformations where folic acid is thought to reduce the risk. It is therefore possible that the teratogenic effect of anticonvulsants is associated with a different mechanism to that associated with the folate metabolism. As folic acid supplementation is not known to be associated with any harmful effects in women of childbearing age, women taking anticonvulsant therapy are however, still recommended to take 5mg daily.

5.4 Pregnancy exposure registries for anticonvulsants

The increase in risk of major congenital malformations associated with exposure to the older anticonvulsants, the inability to discontinue treatment and the potential for inadvertent exposure during pregnancy, make new anticonvulsant drugs ideal candidates for monitoring by pregnancy exposure registries. The introduction of the anticonvulsant lamotrigine in 1992 saw the creation of the first pregnancy registry for an anticonvulsant exposure.⁽²⁹⁾ This was then followed by several pregnancy exposure registries that were set up in the mid to late 1990s to recruit and capture pregnancies exposed to any anticonvulsant product.^(11, 30-32)

Over time these pregnancy exposure registries have captured a large amount of data on anticonvulsant exposure during pregnancy and in some cases have confirmed and provided further supporting evidence for the increased risks associated with first trimester exposure to the older anticonvulsants.⁽¹³⁾ For the newer anticonvulsants, however, many of these registries are still limited in terms of the number of first trimester exposures enrolled for which they have informative outcome data. When evaluating anticonvulsant exposure, identifying a sufficient number of exposures is further restricted by the fact that around 20%

of women take more than one anticonvulsant as polytherapy, making it impossible in these cases to determine any causal association between an individual drug and congenital malformation risk. The variations in data collection methods, in terms of inclusion and exclusion criteria, timing of enrolment, outcome definitions, duration of follow-up and the fact that some pregnancies may be reported to multiple registries means that it is considered inappropriate to pool data from the different registries to increase sample sizes.⁽³³⁾

In 2006 the North American Antiepileptic Drug Registry generated a signal associated with first trimester monotherapy exposure to lamotrigine and a 10-fold increased risk of isolated oral clefts (RR 10.4; CI₉₅ 4.5-24.9). Following this signal, data from other pregnancy registries were analysed and studies in the literature reporting on first trimester exposures to lamotrigine were also reviewed. None of these data sources, however, could provide sufficient evidence to either confirm or refute the association.⁽³⁴⁾ As a result, GlaxoSmithKline, the manufacturer of lamotrigine, decided to continue surveillance with the different pregnancy registries whilst also conducting a case-control study using data collected by the EUROCAT network; a network of population based congenital anomaly registers which cover around 25% of all births in Europe.^(34, 35) This case-control study did not find evidence to confirm the large increase in risk identified by the North American Registry. Given the small numbers, however, it was not possible to rule out anything smaller than a three-fold increase in risk and therefore continued surveillance was recommended.⁽³⁴⁾ It is not clear why the North American Registry came up with such a large association that could not be reproduced elsewhere. Oral clefts can run in families and it is possible that the voluntary nature of enrolment to the registry means women who have already given birth to a child with an oral cleft are more likely to choose to enrol for a subsequent pregnancy. The authors do mention that the prevalence of isolated oral clefts and particularly cleft palate in the population they used as a comparator group was lower than many other data sources. They did demonstrate, however, that when applying these higher comparator figures, although the risk is lowered, the significant increase in risk

did not go away. It will therefore be interesting to see what findings future studies with the advantage of larger sample sizes produce.

In recent years two pregnancy registries have reported findings suggestive of an increased risk of major congenital malformations associated with first trimester exposure to topiramate.^(36, 37) In 2008, the UK Epilepsy and Pregnancy Register reported a major congenital malformation prevalence of 4.8% and 11.2% following mono and polytherapy exposure to topiramate respectively.⁽³⁶⁾ It was also reported that the rate of oral clefts for all topiramate exposed pregnancies (N=203) was eleven times the background rate (2.2%; 95% CI 0.9 to 5.6%). This study also reported 4 cases of hypospadias within 78 male live births (5.1%; 95% CI 0.2 to 10.1), although only 2 of the malformations were severe enough to be classed as major and only one of those was exposed to topiramate as monotherapy.⁽³⁶⁾ Following these findings a study was conducted using data collected by the North American Antiepileptic Drug Pregnancy Register.⁽³⁷⁾ This study identified 289 first trimester monotherapy exposures and it identified a higher prevalence of major malformations following topiramate exposure compared to the unexposed control group of friends and family members (3.8 v 1.3%). It also identified an increased risk of isolated cleft lip compared with the expected background prevalence (0.69 v 0.07%). The results of the latter study informed the FDA's decision to change the safety categorisation of topiramate use during pregnancy from category C to category D (Table 5.1). Both studies, however, concluded that the findings were based on small numbers and further studies were required to determine whether there is evidence of a true causal association.

Table 5.1 Categories used by the US Food and Drug Administration to summarise the safety of medicines used during pregnancy based on the available evidence⁽³⁸⁾

Category	Description of evidence
A	Adequate and well-controlled studies have failed to demonstrate a risk to the fetus in the first trimester of pregnancy (and there is no evidence of risk in later trimesters).
B	Animal reproduction studies have failed to demonstrate a risk to the fetus and there are no adequate and well-controlled studies in pregnant women.
C	Animal reproduction studies have shown an adverse effect on the fetus and there are no adequate and well-controlled studies in humans, but potential benefits may warrant use of the drug in pregnant women
D	There is positive evidence of human foetal risk based on adverse reaction data from investigational or marketing experience or studies in humans, but potential benefits may warrant use of the drug in pregnant women despite potential risks.
X	Studies in animals or humans have demonstrated foetal abnormalities and/or there is positive evidence of human foetal risk based on adverse reaction data from investigational or marketing experience, and the risks involved in use of the drug in pregnant women clearly outweigh potential benefits.

The lamotrigine and topiramate examples demonstrate that pregnancy exposure registries are effective at generating potential safety signals but not every signal is evidence necessarily, of a causal association between exposure and outcome. When evaluating the risks associated with specific defect types, however, sample size within these registries can be a major limitation both in terms of the time taken to generate a signal and the statistical power available to determine an increase in risk with any level of certainty. Over time the increase in the number of pregnancies exposed to newer anticonvulsants, or any new product, is inevitably going to lead to an increase in the number of signals generated in relation to a potential major congenital malformation risk. It is important that potential signals are identified as soon as possible following a product's introduction on the market and that any signals are confirmed or refuted promptly to reduce unnecessary concern and confusion. It is therefore important that all potential data sources that could potentially play a role in achieving this are comprehensively evaluated. The amount of comparison data available on anticonvulsant exposures combined with the chronic nature of the disease and most repeat anticonvulsant prescribing taking place in UK primary care makes anticonvulsants a suitable case study for investigating the potential of the

General Practice Research Database as a tool for drug safety in pregnancy research.

5.5 References

1. Oxford concise colour medical dictionary. 2nd ed. Martin EA, editor. Oxford: Oxford University Press; 1998.
2. Duncan JS, Sander JW, Sisodiya SM, et al. Adult epilepsy. *Lancet*. 367(9516):1087-100.
3. Forsgren L, Beghi E, Öun A, et al. The epidemiology of epilepsy in Europe – a systematic review. *Eur J Neurol*. 2005;12(4):245-53.
4. Craig JJ, Hunt SJ. Treating women with juvenile myoclonic epilepsy. *Practical Neurology*. 2009;9(5):268-77.
5. Olafsson E, Hallgrimsson JT, Hauser WA, et al. Pregnancies of Women with Epilepsy: A Population-Based Study in Iceland. *Epilepsia*. 1998;39(8):887-92.
6. Fairgrieve SD, Jackson M, Jonas P, et al. Population based, prospective study of the care of women with epilepsy in pregnancy. *BMJ*. 2000;321(7262):674-5.
7. Scottish Obstetric Guidelines and Audit Project. The Management of Pregnancy in Women with Epilepsy - A clinical Practice Guideline for Professionals Involved in Maternity Care in Scotland. (1997).
8. Quality and Outcomes Framework guidance for GMS contract 2011/12. Delivering investment in general practice. (2011) Available from: http://www.nhsemployers.org/Aboutus/Publications/Documents/QOF_guidance_GMS_contract_2011_12.pdf [Accessed on 25.01.2012].
9. Fried S, Kozer E, Nulman I, et al. Malformation rates in children of women with untreated epilepsy: a meta-analysis. *Drug Saf*. 2004;27(3):197-202.
10. Wyszynski DF. Pregnancy exposure registries: Academic opportunities and industry responsibility. *Birth Defects Res A Clin Mol Teratol*. 2009;85:93-101.
11. Morrow J, Russell A, Guthrie E, et al. Malformation risks of antiepileptic drugs in pregnancy: a prospective study from the UK Epilepsy and Pregnancy Register. *J Neurol Neurosurg Psychiatry*. 2006;77(2):193-8.
12. Meador K, Reynolds MW, Crean S, et al. Pregnancy outcomes in women with epilepsy: A systematic review and meta-analysis of published pregnancy registries and cohorts. *Epilepsy Research*. 2008;81(1):1-13.
13. Wyszynski DF, Nambisan M, Surve T, et al. Increased rate of major malformations in offspring exposed to valproate during pregnancy. *Neurology*. 2005;64(6):961-5.
14. Tomson T, Battino D, Bonizzoni E, et al. Dose-dependent risk of malformations with antiepileptic drugs: an analysis of data from the EURAP epilepsy and pregnancy registry. *Lancet Neurol*. 2011;10(7):609-17.
15. Meador KJ, Baker GA, Browning N, et al. Cognitive function at 3 years of age after fetal exposure to antiepileptic drugs. *N Engl J Med*. 2009;360(16):1597-605.
16. Cummings C, Stewart M, Stevenson M, et al. Neurodevelopment of children exposed in utero to lamotrigine, sodium valproate and carbamazepine. *Arch Dis Child*. 2011.

17. International League Against Epilepsy. Guidelines for the Care of Women of Childbearing Age with Epilepsy. *Epilepsia*. 1993;34(4):588-9.
18. Montouris G. Importance of monotherapy in women across the reproductive lifecycle. *Neurology*. 2007;69:S10-S6.
19. Group ES. Seizure control and treatment in pregnancy: observations from the EURAP epilepsy pregnancy register. *Neurology*. 2006;66(3):354-60.
20. Lindhout D, Omtzigt JGC. Teratogenic Effects of Antiepileptic Drugs: Implications for the Management of Epilepsy in Women of Childbearing Age. *Epilepsia*. 1994;35:S19-S28.
21. Rosa FW. Spina Bifida in Infants of Women Treated with Carbamazepine during Pregnancy. *N Engl J Med*. 1991;324(10):674-7.
22. MRC Vitamin Study Research Group. Prevention of neural tube defects: Results of the Medical Research Council Vitamin Study. *Lancet*. 1991;338(8760):131-7.
23. Czeizel AE, Dudas I. Prevention of the first occurrence of neural-tube defects by periconceptional vitamin supplementation. *N Engl J Med*. 1992;327(26):1832-5.
24. Hernandez-Diaz S, Werler MM, Walker AM, et al. Folic Acid Antagonists during Pregnancy and the Risk of Birth Defects. *N Engl J Med*. 2000;343(22):1608-14.
25. National Institute of Clinical Excellence. The epilepsies - the diagnosis and management of the epilepsies in adults and children in primary and secondary care. 2004 Available from <http://publications.nice.org.uk/the-epilepsies-the-diagnosis-and-management-of-the-epilepsies-in-adults-and-children-in-primary-and-cg137>. [Accessed on 16.11.2011].
26. Morrow JI, Hunt SJ, Russell AJ, et al. Folic acid use and major congenital malformations in offspring of women with epilepsy: a prospective study from the UK Epilepsy and Pregnancy Register. *J Neurol Neurosurg Psychiatry*. 2009;80(5):506-11.
27. Jentink J, Bakker MK, Nijenhuis CM, et al. Does folic acid use decrease the risk for spina bifida after in utero exposure to valproic acid? *Pharmacoepidemiol Drug Saf*. 2010;19(8):803-7.
28. Craig J, Morrison PJ, Morrow J, et al. Failure of periconceptual folic acid to prevent a neural tube defect in the offspring of a mother taking sodium valproate. *Seizure*. 1999;8:253-4.
29. Cunnington MC. The International Lamotrigine pregnancy registry update for the epilepsy foundation. *Epilepsia*. 2004;45(11):1468.
30. Vajda FJ, Hitchcock A, Graham J, et al. Foetal malformations and seizure control: 52 months data of the Australian Pregnancy Registry. *Eur J Neurol*. 2006;13(6):645-54.
31. Tomson T, Battino D, Bonizzoni E, et al. EURAP: an international registry of antiepileptic drugs and pregnancy. *Epilepsia*. 2004 Nov;45(11):1463-4.
32. Holmes LB, Wyszynski DF. North American antiepileptic drug pregnancy registry. *Epilepsia*. 2004;45(11):1465.
33. Tomson T, Battino D, French J, et al. Antiepileptic drug exposure and major congenital malformations: The role of pregnancy registries. *Epilepsy Behav*. 2007;11(3):277-82.
34. Dolk H, Jentink J, Loane M, et al. Does lamotrigine use in pregnancy increase orofacial cleft risk relative to other malformations? *Neurology*. 2008;71:714-22.
35. Dolk H. EUROCAT: 25 years of European surveillance of congenital anomalies. *Arch Dis Child Fetal Neonatal Ed*. 2005;90:F355-F8.
36. Hunt S, Russell A, Smithson WH, et al. Topiramate in pregnancy: preliminary experience from the UK Epilepsy and Pregnancy Register. *Neurology*. 2008;71(4):272-6.
37. Hernandez-Diaz S, Mittendorf R, Holmes LB. Comparative safety of topiramate during pregnancy. *Birth Defects Res A Clin Mol Teratol*. 2010;88(5):408.

38. University of Washington Medicine. FDA pregnancy categories. UW Medicine Drug Formulary. Available from:
<http://depts.washington.edu/druginfo/Formulary/Pregnancy.pdf> [Accessed on 23.09.2011]

Chapter 6

Identifying major congenital malformations in the UK General Practice Research Database

This article has been reproduced with permission from Adis, a Wolters Kluwer business (see Charlton RA, Weil JG, Cunnington MC, *et al.* Identifying major congenital malformations in the UK General Practice Research Database (GPRD): a study reporting on the sensitivity and added value of photocopied medical records and free text in the GPRD. *Drug Saf* 2010; 33 (9): 741-50.) for the final published version).[©] Adis Data Information BV 2010. All rights reserved.

6.1 Introduction

The UK General Practice Research Database (GPRD) is the world's largest computerized database of anonymised, longitudinal medical records from primary care,⁽¹⁾ and has the potential, in some circumstances, to aid research into the safety of medicines used during pregnancy.⁽²⁾ Work has begun to validate the recording of congenital malformations in the GPRD, although to our knowledge thus far it has been limited to a selection of cardiovascular defects⁽³⁾ and neural tube defects.⁽⁴⁾

This paper forms part of a broader study to assess the potential of the GPRD to act as a pregnancy registry system focusing on a cohort of women with a diagnosis of epilepsy, seizure or convulsion. Here we describe the methodology used for the surveillance of the range of congenital malformations that would be required if the GPRD is to be used to monitor drug safety in pregnancy. We report the extent to which diagnoses identified in the GPRD could be confirmed, rejected or made more specific by using the photocopied paper medical records or where this was not possible any data entries consisting of uncoded comments, so-called 'free text,' in the GPRD, which are not routinely available for research purposes.

6.2 Methods

6.2.1 Data source

The GPRD contains over 44 million person-years of data and currently captures data from ~4 million active patients (~7% of the UK population).⁽⁵⁾ Virtually all prescriptions, non-drug interventions and referrals issued by general practitioners (GPs) are recorded in the database, as are medical diagnoses, including those relating to pregnancy. The database is managed by the GPRD division at the Medicines and Healthcare products Regulatory Authority (MHRA), who also provide a number of services to allow verification of events and diagnoses identified from the computerized records. For patients still registered at the GP practice it is possible to request anonymised photocopies of the patient's full paper medical record, enabling access to all referrals and out-patient letters and correspondence from consultants and specialists. For all

patients, regardless of whether they are still registered with the GP practice, it is possible to request the information recorded by GPs as free text in the electronic GPRD record. GPs have the option of recording free text comments, i.e. uncoded additional information relating to symptoms, more detailed diagnoses, test results, etc., each time they record a medical code within the database. It is recommended that these verification services are used to assess the sensitivity and specificity of medical codes.⁽¹⁾ All methods of verification are requested via the GPRD group at the MHRA to ensure that both practice and patient confidentiality are maintained.

6.2.2 Study population

Women were eligible for inclusion if they were, or had been, permanently registered at a GP practice regarded by the MHRA to be contributing data up to research standard. Women were included in the analysis of verification of major congenital malformation codes within the child's record if they had a medical record indicating a live pregnancy outcome between 1 January 1990 and 31 December 2006 and they had a code relating to a diagnosis of epilepsy, seizure or convulsion.

Women were excluded from the cohort if they were not 14–49 years of age at the date of delivery and if they did not have a code indicating a pregnancy in the 280 days leading up to the delivery date. In view of requirements for the broader study, women were also excluded if they were not continuously enrolled in the GPRD for the 4 months before the estimated date of the last menstrual period or if the diagnosis of epilepsy, seizure or convulsion was not before the pregnancy indicator code.

The offspring of women meeting all inclusion criteria were identified where possible (based on having the same family and GP practice numbers and the child's year and month of birth being equal to the mother's year and month of delivery). Mother-baby pairs were included if the linked child was registered and present in the GPRD 3 months after the mother's pregnancy outcome date. If the child had been registered and died before reaching 3 months of age the mother-

baby pair was still included and the need to be registered 3 months after the pregnancy outcome date was no longer required. Women were entitled to contribute more than one mother-baby pair to the study and each unique offspring was considered separately.

6.2.3 Identification of major congenital malformations

To identify individuals with a major congenital malformation, search terms for all congenital malformations were created based on those listed in the International Classification of Diseases 9th edition⁽¹¹⁾ ‘congenital anomalies’ chapter (ICD-9 codes 740–759). These search terms contained ‘wild cards’ to allow for variations in terms, e.g. ‘hydroceph*’ to account for ‘hydroceph-aly’ and ‘hydroceph-alus’. All medical codes containing these search terms were then selected plus any relevant codes identified by being in their hierarchical vicinity. Children of women in the cohort were then identified in the GPRD if they had any of these codes within their medical record.

As pregnancy registries are primarily concerned with major malformations, the malformations identified were categorized as major or minor based on the information available through the Read and OXMIS codes used in the GPRD, photocopied medical records and free text. This categorization was based on the classification used by the European network of population-based registers for the epidemiological surveillance of congenital anomalies (EUROCAT). Minor defects and malformations associated with prematurity, when isolated, are excluded from EUROCAT reports. In addition, there are some malformations (e.g. hypospadias, hydronephrosis, talipes, syndactyly) that are only classified as major if certain criteria are met.^(6,7)

All congenital malformation codes identified within the cohort were reviewed independently by two of the authors (RC, JW). Those relating to a major malformation and those relating to a malformation that could be classed as major, if specific criteria were met, were selected for verification.

6.2.4 Verification of congenital malformations

Verification of congenital malformations was carried out by requesting a photocopy of the child's entire medical record for children still registered with the practice. These were then reviewed to ensure that the congenital malformation recorded in the computerized medical record was a true malformation and had been correctly coded. Secondly, for those malformations that needed to meet specific inclusion criteria in order to be classified as major, information relating to this was identified, where present, enabling the true number of major malformations to be determined. Photocopied medical records were requested via the MHRA, who sent out an initial request and three subsequent reminders to GPs at fortnightly intervals. For children who had transferred out of the practice and for those where the GP did not return photocopied records, all the information recorded in the free text within the child's entire electronic GPRD medical record was obtained and reviewed.⁴ One example recorded alongside a code for an oral cleft included: "Right cleft lip – admit ***, repair cleft lip under GA, Hare lip palate intact." All information in the photocopied medical records and free text was anonymised by the MHRA before being transferred to the investigators.

6.2.5 Analyses

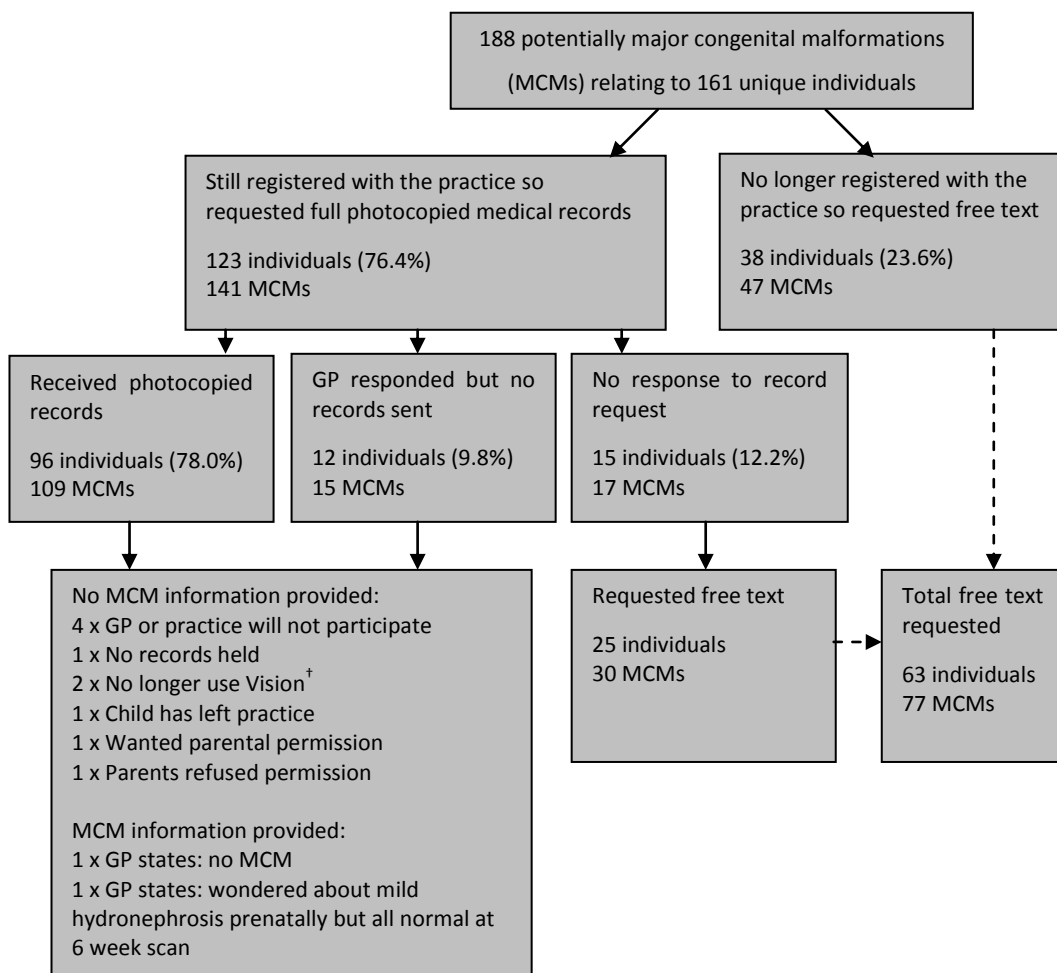
The percentage of diagnoses identified from the computerized records that could be confirmed as being the congenital malformation recorded was calculated for the photocopied medical records and free text verification methods separately and combined. Of those malformations that were confirmed, the percentage where sufficient information was available to classify as a major or minor congenital malformation was then calculated both overall and for different subgroups of congenital malformations. As we did not request photocopied medical records or free text for those individuals who were not identified as having a major congenital malformation, it was not possible to calculate the specificity of the computerized records.

⁴ It is not possible to obtain photocopied records for patients who have transferred out of the practice.

6.3 Results

6.3.1 Verifying the presence of a congenital malformation

The study population consisted of 3869 live births within this cohort of women with a record of epilepsy, seizure or convulsion any time before the pregnancy. 188 potentially major congenital malformations recorded at any age were identified relating to 161 unique individuals. Photocopied medical records were requested for the 123 individuals (76.4%) still registered with the practice. Figure 6.1 shows the response to the photocopied record requests, with 96 records (78.0%) being returned, relating to 109 unique malformations. In addition, the GPs of 12 individuals (9.8%) replied to explain why they were not enclosing the photocopied records.



[†] The GP practice software that feeds into the GPRD; MCM = major congenital malformation

Figure 6.1 Methods used for verifying malformations

For 15 individuals (12.2%) the GP did not respond to the record request and therefore free text was obtained. Free text was also obtained for the ten cases where the GP responded but did not provide additional information on the potential major congenital malformation and for the 38 individuals no longer registered with the practice. In total, free text was therefore obtained for 63 individuals (39.1%) representing 77 major congenital malformations. Figures 6.2 and 6.3 along with Table 6.1 summarise the percentage of malformations that could be confirmed (i.e. verified) or refuted as a congenital malformation, and the percentage of those confirmed that could be further classified as being major or minor.

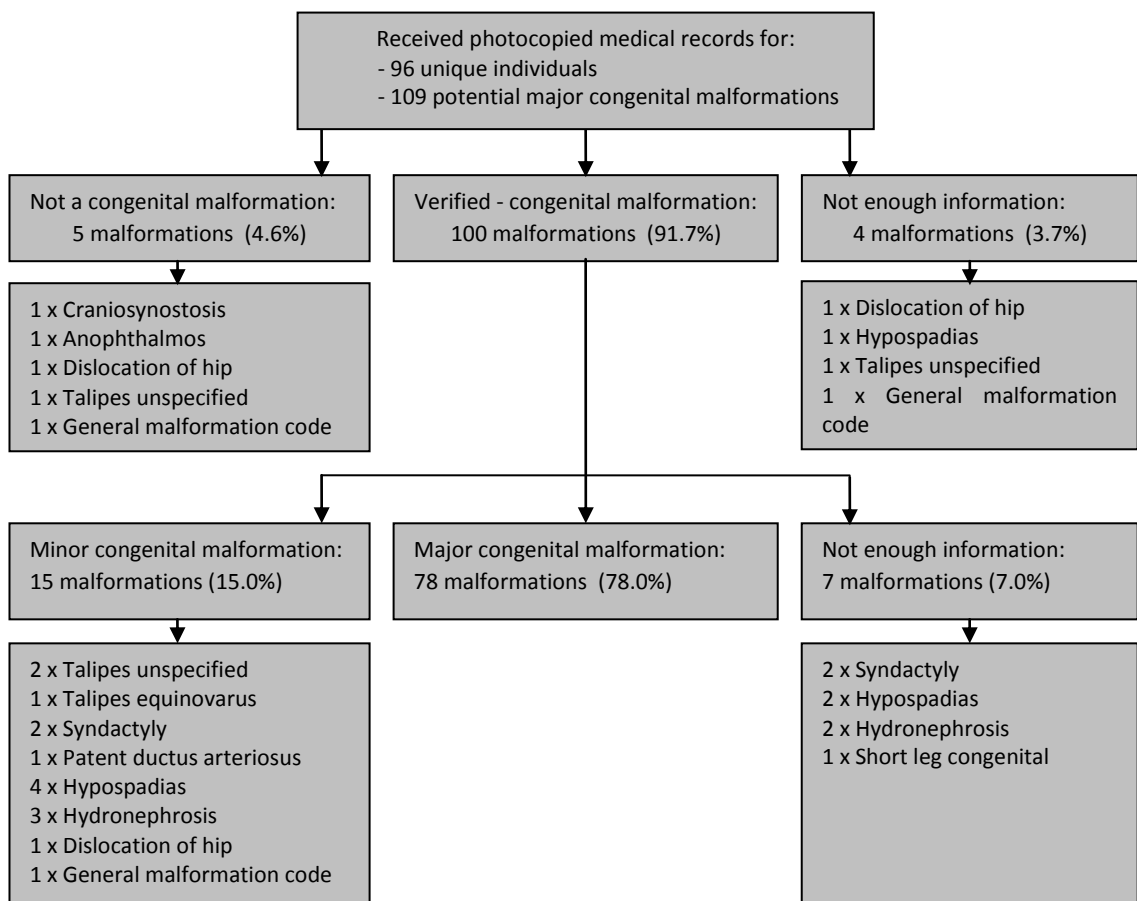


Figure 6.2 Percentage of congenital malformations that could be verified using photocopied records and of those the percentage that could be classified as major or minor. Major/minor classification was based on that used by EUROCAT.

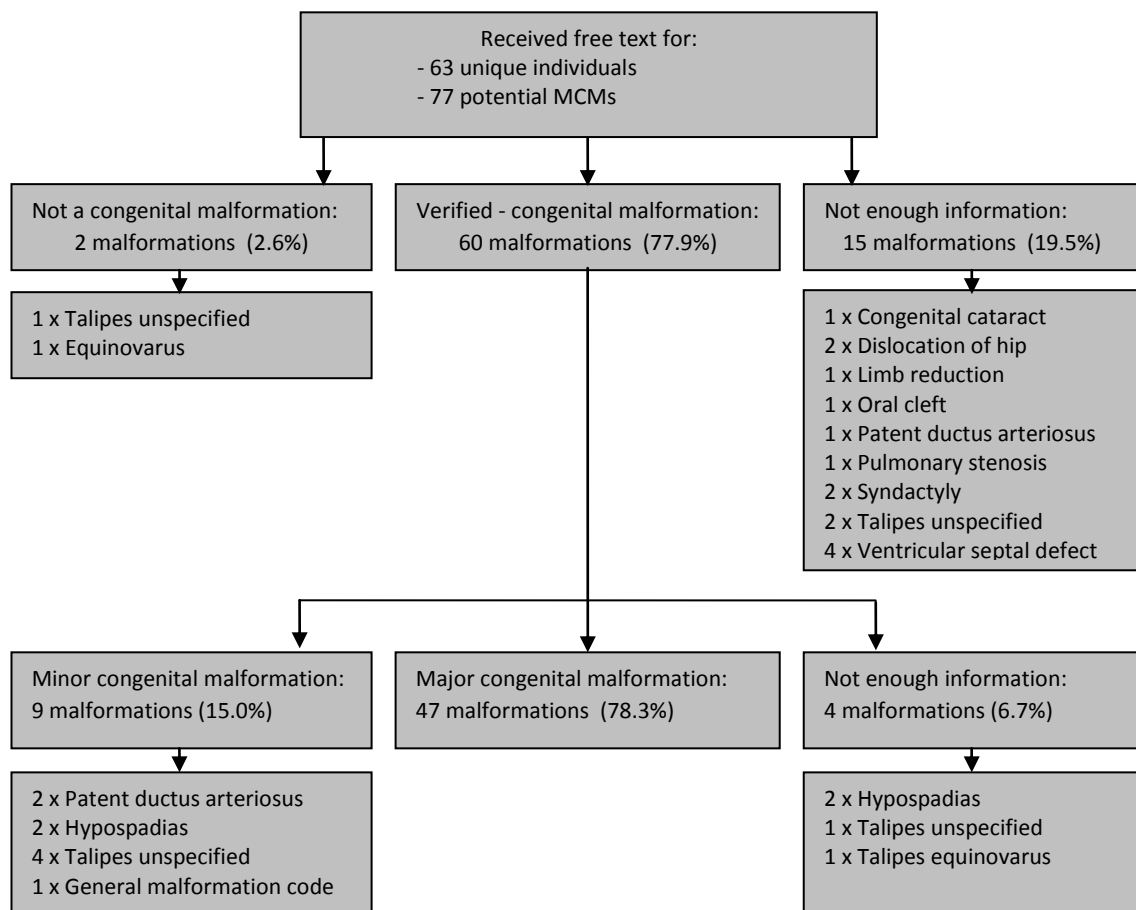


Figure 6.3 Percentage of congenital malformations that could be verified using free text and of those the percentage that could be classified as major or minor. Major/minor classification was based on that used by EUROCAT.

Using a combination of photocopied medical records and free text it was possible to verify 160 malformations (85.1%) as the malformation indicated by the computerized records. For those where photocopied records were available, the percentage verified was 91.7% (100/109) and for those where the free text was used it was possible to verify 77.9% of malformations (60/77). Nine cases (4.8%) in total were found not to have the malformation recorded; these included seven cases of diagnostic suspicion that were not confirmed on investigation and where computer records were not updated, one case where the GP wrote in response to receiving the request for photocopied medical records to state that the infant did not have a major congenital malformation and one case where an incorrect code had been recorded. Using photocopied medical records it was possible to either confirm or refute the presence of a malformation in 96.3% of cases (105/109) and in 80.5% of cases (62/77) using the free text.

Table 6.1 Percentage of congenital malformations that could be verified and, of those, the percentage that could be classified as major or minor

Verification method	Number of malformations	Outcome of verification of congenital malformations identified by medical codes in the GPRD [n (%)]			Verified congenital malformations that could be classified as major or minor (n = 160) [n (%)]		
		yes (a malformation)	no (refute, incorrectly coded)	not enough information	major	minor	not enough information
Photocopied medical records ^b	109	100 (91.7)	5 (4.6)	4 (3.7)	78 (78.0)	15 (15.0)	7 (7.0)
Free text ^c	77	60 (77.9)	2 (2.6)	15 (19.5)	47 (78.3)	9 (15.0)	4 (6.7)
Both methods combined ^a	188 ^a	160 (85.1)	9 ^a (4.8)	19 (10.1)	125 (78.1)	24 (15.0)	11 (6.9)

^a Includes the two cases where the general practitioner wrote to say that there was no malformation but did not send photocopied medical records and free text was not requested.

^b Anonymised photocopies of the patient's full paper medical record, enabling access to all referrals and out-patient letters and correspondence from consultants and specialists.

^c Anonymised uncoded additional information, i.e. relating to symptoms, more detailed diagnoses, test results, etc. recorded by the general practitioner in the patient's computerised medical record.

GPRD = General Practice Research Database.

Table 6.2 demonstrates that the verification of the presence of a congenital malformation varied by malformation class. It was possible to verify all CNS, digestive system, chromosomal and foetal valproate syndrome anomalies. Within the majority of other malformation classes, with the exception of talipes, congenital dislocation or dysplasia of the hip and anomalies of the eye, verification was found to be over 80.0%.

6.3.2 Classifying as a major or minor congenital malformation

Table 6.1 shows that of the 160 verified congenital malformations, it was possible to classify 93.1% (149) as being major or minor. This percentage was found to be the same for those cases reviewed by photocopied medical records and those reviewed using free text. For 125 cases (78.1% of those verified), review of the photocopied medical records or free text resulted in the malformation identified on the computerized records being classified as major, whilst 24 cases (15.0%) were classified as minor. For the remaining 6.9% (11/160) of verified malformations there was insufficient information to classify them as being major or minor. Of the 188 congenital malformations initially identified, 125 (66.5%) could be both verified as a malformation and classified as being major, which is what would be of primary interest to pregnancy registries.

The percentage of verified congenital malformations that could be classified as major was found to vary by malformation class (Table 6.2). For many classes of malformation the percentage that could be classified as major was $\geq 80.0\%$. This was, however, lower for those classes that included malformations needing to meet specific inclusion criteria (e.g. hydronephrosis, hypospadias, syndactyly, talipes, patent ductus arteriosus) to be classified as being major (e.g. talipes – where cases of a postural origin were excluded from the major malformation count).

Table 6.2 The number of congenital malformations that could be verified, and of those the number that could be classified as major, by malformation class

Malformation class	Number of malformations	Outcome of verification of congenital malformations identified by medical codes in the GPRD (n)			Of the verified congenital malformations the number that could be classified as major or minor (n)		
		Yes, a malformation	No, refute, incorrectly coded	Not enough information	Major	Minor	Not enough information
Central nervous system	5	5	0	0	5	0	0
Congenital heart defects	50	44	0	6	41	3	0
Orofacial clefts	7	6	0	1	6	0	0
Eye	4	2	1	1	2	0	0
Digestive system	6	6	0	0	6	0	0
Internal urogenital system	23	22	1	0	17	3	2
Hypospadias	28	27	0	1	17	6	4
Talipes	18	12	3	3	4	6	2
Hip dislocation/dysplasia	16	11	2	3	10	1	0
Poly-/syndactyly	10	8	0	2	4	2	2
Limb reduction	6	5	0	1	4	0	1
Musculoskeletal	1	0	1	0	NA	NA	NA
Chromosomal	1	1	0	0	1	NA	NA
Foetal valproate syndrome	4	4	0	0	4	0	0
Other	9	7	1	1	4	3	0
Total	188	160	9	19	125	24	11

NA = not applicable; GPRD = General Practice Research Database.

6.4 Discussion

This study indicates that a wide range of congenital malformations can be identified reliably using the GPRD. The presence or absence of a malformation, regardless of whether major or minor, was confirmed in 85.1% (160/188) of those identified using the computerized records. In infants for whom photocopied medical records were received, it was possible to confirm or refute 96.3% of recorded malformations. The overall verification rate was lowered by the review of some cases being limited to information in the free text, but this resulted largely from the study covering a retrospective 16-year time period. Had the study been carried out in real-time following a drug being launched on the market, or had a review been carried out at frequent intervals, then photocopied medical records would have been available for a larger proportion of individuals, probably resulting in higher verification rates.

At 78.1%, the overall verification and classification of major malformations within this cohort of individuals in the GPRD compares favourably to that found by Cooper et al.⁽¹⁰⁾ using Tennessee Medicaid computerized records. They found that 67.7% of all major congenital malformations identified by birth certificates or patient claims could be confirmed by medical record review and that medical records were available for 98.9% of individuals. Two studies have identified major congenital malformations using the Ingenix Research Data Mart (which contains medical and pharmacy claims data from United Healthcare affiliated health plans) and verified them by medical record abstraction.^(8,9) Only one study reported on the availability of medical records and that they were available for 86% of potential major congenital malformation cases.⁽⁸⁾ Both studies reported only the number of confirmed major congenital malformations and did not report the proportion of those identified via the computerized records that could be verified, so direct comparison with our study was not possible.

Verification rates in the GPRD, as with the Tennessee Medicaid records, appeared to vary by the type of defect recorded. Malformations that were compatible with life and required medical intervention or monitoring could be

verified relatively easily. For malformations that could be defined as being either major or minor depending on anatomical details regarding the malformation (e.g. hypospadias, hydronephrosis, syndactyly), as well as for malformations that could resolve spontaneously, it proved more difficult to obtain the necessary detail. For instance, the six congenital heart defects where there was insufficient information to confirm the diagnosis consisted of four ventricular septal defects, one patent ductus arteriosus and one case of pulmonary stenosis. We were unable to obtain the photocopied records for any of these cases and no additional information had been recorded in the free text. It is possible that these were less serious, largely asymptomatic defects that did not prompt the GP to record additional information in the free text. For six hypospadias cases (21.4%) there was insufficient information available to classify as major or minor. The EUROCAT guidelines, however, changed in 2005 to include the reporting of all cases of hypospadias as a major malformation and therefore the need for additional detail to distinguish between major and minor types is no longer an issue.

Compared with the photocopied medical records, the free text was less likely to provide sufficient information to confirm or refute the presence of a malformation. However, if there was evidence that the malformation was present, then using the free text enabled the same percentage of malformations to be classified as major or minor as with the photocopied medical records. With more GP practices scanning patient letters into the free text it is possible that the level of detail recorded in the free text has increased over time.

Within this study we chose to request and review full photocopied medical records in order to obtain as much information as possible about individuals still registered on the database. Previous verification studies of major congenital malformations, however, involved using a medical record abstraction questionnaire, which has also been found to be successful. Wurst *et al.*,⁽³⁾ when validating three specific types of congenital heart defect, reported a 94% response rate from GPs to the questionnaire and identified a positive predictive

value $\geq 90\%$ for each defect type. Devine *et al.*⁽⁴⁾ created an algorithm to identify cases of neural tube defects from the computerized records and then sent questionnaires to verify the diagnoses. This study reported a 76.0% response rate to the questionnaire, although a small number were sent to GPs where the patient concerned had already transferred out of the practice. The overall positive predictive value of the algorithm for neural tube defects was 71%, ranging from 47% for spina bifida to 83% for encephalocele. Questionnaires are, therefore, an additional tool that could be used to verify diagnoses, although when attempting to verify a wide range of malformations the photocopied medical records have the advantage of not requiring the design of a large number of malformation-specific questionnaires.

This study has demonstrated that in a large number of cases, when a malformation is identified in the GPRD via the computerized medical records, the malformation is likely to exist. There will, however, be a small percentage of identified cases (~5%) that need to be excluded due to being incorrectly coded or diagnostically ruled out. Therefore, making selected information recorded in the free text such as 'excluded' or 'ruled out' routinely available to researchers would be beneficial. For a slightly larger proportion, the malformation identified is present but it may be in a minor form. However, for those who are interested in aetiology, the inclusion of these cases may well be of value.

6.5 Conclusion

Postmarketing surveillance programmes are essential to monitor drug safety in pregnancy, and require a data source that can reliably capture cases of congenital malformations. This study has demonstrated the GPRD can be used to ascertain a wide range of congenital malformations. Photocopied medical records and to a lesser extent free text have proven valuable sources in carrying out verification of malformations identified through the computerized records. For more severe malformations and those that are easily externally visible, the computerized records were found to be reliable; verification is, however, still recommended, especially for those malformations that can occur with different

levels of severity. Further work might assess the accuracy and completeness of the computerized GPRD malformation codes in a different study population and in relation to terminations of pregnancy.

6.6 References

1. Wood L, Martinez C. The General Practice Research Database: role in pharmacovigilance. *Drug Saf* 2004;27(12):871-81
2. Charlton RA, Cunningham MC, de Vries CS, et al. Data resources for investigating drug exposure during pregnancy and associated outcomes: the General Practice Research Database (GPRD) as an alternative to pregnancy registries. *Drug Saf* 2008;31(1):39-51
3. Wurst KE, Ephross SA, Loehr J, et al. The utility of the General Practice Research Database to examine selected congenital heart defects: a validation study. *Pharmacoepidemiol Drug Saf* 2007;16:867-77
4. Devine S, West SL, Andrews E, et al. Validation of neural tube defects in the full-featured General Practice Research Database. *Pharmacoepidemiol Drug Saf* 2008;17:434-44
5. MHRA. The General Practice Research Database, 2009 [online]. Available from URL <http://www.gprd.com/home/default.asp> [Accessed 01.02.2010]
6. EUROCAT. List of minor anomalies for exclusion up to birth year 2004 [online]. Available from: <http://www.eurocat-network.eu/content/EUROCAT-Old-List-Minor-Anomalies.pdf> [Accessed 26.01.2010]
7. EUROCAT. Minor anomalies for exclusion (chapter 3.2, guide 1.3), 2005 [online]. Available from: <http://www.eurocat-network.eu/content/EUROCAT-Guide-1.3.pdf> [Accessed 26.01.2010]
8. Cole JA, Modell JG, Haight BR, et al. Bupropion in pregnancy and the prevalence of congenital malformations. *Pharmacoepidemiol Drug Saf* 2007;16:474-84
9. Cole JA, Ephross SA, Cosmatos IS, et al. Paroxetine in the first trimester and the prevalence of congenital malformations. *Pharmacoepidemiol Drug Saf* 2007;16:1075-85
10. Cooper WO, Hernandez-Diaz S, Gideon P, et al. Positive predictive value of computerised records for major congenital malformations. *Pharmacoepidemiol Drug Saf* 2008;17:455-60
11. Hart AC, Hopkins CA, editors. ICD-9-CM professional for hospitals. Volumes 1, 2 and 3. International classification of diseases 9th revision, clinical modification 6th ed. Reston (VA): Ingenix Inc. 2003, St. Anthony Publishing/ Medicode, 2003

Chapter 7

Comparing the General Practice Research Database and the UK Epilepsy and Pregnancy Register as tools for postmarketing teratogen surveillance

This article has been reproduced with permission from Adis, a Wolters Kluwer business (see Charlton RA, Weil JG, Cunnington MC, *et al.* Comparing the General Practice Research Database and the UK Epilepsy and Pregnancy Register as Tools for Postmarketing Teratogen Surveillance: Anticonvulsants and the Risk of Major Congenital Malformations. *Drug Saf* 2011;34(2):157-171.) for the final published version).[©] Adis Data Information BV 2011. All rights reserved.

7.1 Introduction

Prescription medications are commonly used by women of childbearing age.^(1,2) With an estimated 30–50% of pregnancies being unplanned^(3,4) and some medical conditions (e.g. epilepsy and depression) making it inadvisable to stop treatment, there is the potential for women to be exposed to medications during the first trimester of pregnancy, which is the critical time period for organ and tissue development. Usually pregnant women are excluded from clinical trial programmes,⁽⁵⁾ and consequently the safety of medicine use during pregnancy and its impact on the risk of congenital malformations cannot be fully assessed until the drug has been marketed.

Pregnancy registries have been used commonly over the past 2 decades to monitor the safety of a new product on the market. Registries aim to detect any substantial increase in the risk of major congenital malformations (MCMs), which are generally defined as those that are life threatening, require major surgery or result in the child having a considerable disability.⁽⁶⁾ Pregnancy registries require primary data collection, which can be time consuming and costly. Given the recent increase in electronic capture of healthcare data for administrative, audit and research purposes, in some circumstances data obtained during routine clinical practice might be able to address the same questions as those explored using pregnancy registries.

The UK General Practice Research Database (GPRD) has been identified as a potential data source for the postmarketing surveillance of drug exposure during pregnancy.⁽⁷⁾ Methods have been developed to identify pregnancies on the database⁽⁸⁻¹⁰⁾ and to link the mothers' medical records with those of the offspring.⁽¹¹⁾ Congenital malformations remain the primary outcome of interest following drug exposure during pregnancy and studies have begun to verify the GPRD (Read/OXMIS) coding system with respect to the identification of some specific defect types (cardiac defects⁽¹²⁾ and neural tube defects⁽¹³⁾).

This study aimed to examine further the potential of the GPRD to serve as a pregnancy registry. Given the presence of a number of pregnancy registries monitoring antiepileptic drug (AED) use, this study aimed to replicate their findings concerning the risk of all MCMs by AED therapy type, and focused the comparison on those reported by the UK Epilepsy and Pregnancy Register.⁽¹⁴⁾ The UK Epilepsy and Pregnancy Register and the GPRD capture similar geographic populations and in both the information on outcome of pregnancy is primarily from the general practitioner (GP), which facilitates a comparison of results. This study also assessed the utility of the GPRD to identify a known teratogenic association, namely the association between first-trimester exposure to valproate and the risk of spina bifida.^(15,16) Replication of this known teratogenic association has also formed part of a validation study by the European network of population-based registers for the epidemiological surveillance of congenital anomalies (EUROCAT) to assess the validity of using EUROCAT data to detect AED-associated risks of specific malformations.⁽¹⁷⁾

7.2 Methods

7.2.1 Data sources

The GPRD is the world's largest computerized database of anonymised, longitudinal medical records from primary care.⁽¹⁸⁾ The GPRD contains over 55 million person-years of data and currently captures approximately 4 million active patients (approximately 7% of the UK population) registered with approximately 500 practices within the UK.⁽¹⁹⁾ Virtually all prescriptions, non-drug interventions and referrals issued by GPs are recorded in the database, as are medical diagnoses, including pregnancy.

The UK Epilepsy and Pregnancy Register is a prospective, observational, registration and follow-up study.⁽¹⁴⁾ Women can either enrol directly into the UK register themselves or they can be enrolled by their healthcare professional (GP, epilepsy specialist, neurologist, etc.). Information on AED exposure and demographic variables are collected from the referring source before the pregnancy outcome is known. Information on pregnancy outcome and the

presence of an MCM is collected approximately 3 months after the estimated date of delivery via a GP questionnaire. Pregnancies can be reported to the UK register retrospectively, after the pregnancy outcome is known, but these are excluded from the standard prevalence estimates to avoid selection bias towards more severe outcomes.

7.2.2 General Practice Research Database maternal study population

The GPRD study followed a retrospective cohort design, and women were eligible for inclusion if they were, or had been, permanently registered at a GP practice considered by the GPRD division at the Medicines and Healthcare products Regulatory Agency (MHRA) to be contributing data up to standard for the purposes of research. The number of eligible pregnancies and MCMs reported in this paper are a subgroup of those reported in an earlier publication by Charlton *et al.*⁽²⁰⁾ More stringent inclusion criteria for epilepsy have been used in the study presented here owing to the need to ensure that the GPRD study population was as comparable as possible to those enrolling in the UK Epilepsy and Pregnancy Register. Sensitivity analyses were carried out to assess the impact on risk estimates of using more stringent inclusion criteria.

In the current study, women were identified as having epilepsy if they had any of the following within their medical record:

- at least two epilepsy diagnosis codes;
- one epilepsy diagnosis code and at least one AED prescription;⁵
- at least two seizure codes (excluding febrile or neonatal seizures) and at least one AED prescription;
- one epilepsy code and at least two seizure codes (excluding febrile or neonatal seizures).

⁵ AEDs included acetazolamide, beclamide, carbamazepine, clobazam, clonazepam, ethosuximide, gabapentin, lamotrigine, levetiracetam, oxcarbazepine, phenobarbital, phenytoin, pregabalin, tiagabine, topiramate, valproate, vigabatrin.

Of those identified with epilepsy, women with a record relating to a pregnancy outcome between 1 January 1990 and 31 December 2006 were identified.

7.2.3 Identifying pregnancy outcomes

For live- and stillbirths (≥ 24 weeks' gestation) the pregnancy outcome date was considered to be the date of the first record of a pregnancy outcome when no additional records were identified in the preceding 90 days. For terminations of pregnancy, the date of the termination was taken as the last recorded termination of pregnancy code, within a 6-week window, as earlier termination of pregnancy codes commonly related to requests and referrals for an elective termination rather than the termination itself. Pregnancy outcomes specifically stating that they were a spontaneous abortion or miscarriage were excluded as they are beyond the scope of a pregnancy registry because of the likelihood of inconsistent identification of defects. Women could contribute more than one pregnancy outcome to the study and each unique pregnancy outcome was considered separately.

Pregnancies were excluded from the cohort if the woman was not aged 14–49 years at the date of the pregnancy outcome and if she did not have any codes indicating a pregnancy (e.g. last menstrual period (LMP), pregnant, positive pregnancy test, antenatal care, etc.) in the 280 days before the pregnancy outcome date. Women were required to be continuously enrolled in the GPRD for the 4 months before the estimated LMP date and throughout the pregnancy to allow reliable assessment of AED exposure. Further pregnancies were excluded if the evidence of epilepsy was not recorded before the first medical record indicative of a pregnancy.

7.2.4 Linking mother-baby pairs

The offspring of women meeting all inclusion criteria described above were identified where possible (based on having the same family and GP practice numbers, and the child's year and month of birth being equal to the mother's year and month of delivery). To enable comparison with the UK register, which

collects outcome data approximately 3 months after the expected delivery date,⁽¹⁴⁾ infants of mother-baby pairs were required to be registered in the GPRD at 3 months of age or to have been registered and died before 3 months of age.

Major congenital malformations (MCMs)

7.2.5 Live births

Medical codes relating to any type of congenital malformation were identified using a list of search terms that was created based on the conditions listed in the 'congenital anomalies' chapter of the *International Classification of Diseases 9th Edition* (ICD-9 codes 740–759). Children of women in the cohort who had ≥ 1 of these codes or relevant codes within their hierarchical vicinity were identified within the GPRD.

As pregnancy registries are primarily concerned with major malformations, the malformations identified were categorized as major or minor. To ensure consistency in terms of classification, malformations were categorized according to the classification used by EUROCAT. The same classification had been used by the UK Epilepsy and Pregnancy Register.⁽¹⁴⁾ Minor defects and malformations associated with prematurity, when isolated, are excluded from EUROCAT reports. There is also a small number of malformations (e.g. hypospadias, hydronephrosis, talipes, syndactyly) that are only classified as major if certain criteria are met.^(21,22) All congenital malformation codes identified within the cohort were reviewed independently by two of the authors.

Malformations identified via the computerized records were confirmed or refuted, for those still registered with the practice, by scrutinizing a photocopy of the child's anonymised full medical record, enabling access to all referral letters, letters from specialists, hospital discharge reports, etc. For children who had transferred out of the practice and for those where the GP did not return photocopied records, information recorded in the free-text fields of the patients' entire medical record was obtained and reviewed. Each time a GP records a medical code within the GPRD they have the opportunity to record any

additional information in the free-text field. This may include a more detailed description of symptoms, test results or information relating to diagnostic procedures and surgery. All information in the photocopied records and free text was anonymised by the MHRA before being returned to the investigators. The analyses in the remainder of this study include all confirmed malformations that were classified as major and all confirmed malformations where there was insufficient information to classify them as minor. For the purpose of this study, chromosomal defects, congenital malformations known to be of genetic origin and malformations where there was clear evidence that the malformation was not drug-induced (e.g. hydrocephalus that was secondary to an intraventricular haemorrhage) were excluded.

7.2.6 Terminations, stillbirths and neonatal deaths

To identify terminations of pregnancy that followed an MCM diagnosis, the free-text comments recorded in the women's electronic medical record during the 2 months before and 4 months after the termination date were obtained. This was because these comments were likely, in some cases, to contain additional information relating to antenatal scans, diagnostic tests and malformation diagnoses. For terminations where the free text did contain information relating to an MCM, this was taken to be sufficient evidence and no further supporting evidence was required. For pregnancies ending in a stillbirth or a neonatal death, free text was obtained for the 2 months before and 6 months after the event. This extended time period was chosen to allow any post-mortem results to be reported back to the GP.

7.2.7 First-trimester exposure

To identify the first trimester (first 13 weeks following the LMP), the start date of a pregnancy (LMP) for a live birth was assumed to be 280 days before the pregnancy outcome date unless there was a record in the woman's medical file indicating that the delivery was pre- or post-mature, in which case the assumed LMP date was adjusted accordingly.

Prescriptions for AEDs, masked to outcome status, were identified and used to establish AED exposure status. Prescriptions were mapped based on the assumption that one prescription could not start until the previous one had finished unless it was being taken as polytherapy or there was clear evidence of drug switching. Where there was evidence of drug switching (i.e. where there were continuous prescriptions for a new drug treatment and no more prescriptions for the original drug), the first drug was deemed to have been discontinued on the date the second one was prescribed. Based on these assumptions, exposure status of study participants was determined for each AED on each day and women were subsequently classified as having been exposed in the first trimester to specific AEDs as well as whether the exposure was mono- or polytherapy.

7.2.8 Terminations of pregnancy

For terminations of pregnancy, gestational age and first-trimester exposure to AEDs was ascertained, for MCM cases, by the manual independent review of the patient's electronic prescription records by two of the authors (RC, CdV), masked to the type of MCM and the pregnancy start and end dates.

7.2.9 Analyses

The prevalence of non-chromosomal MCMs meeting the inclusion criteria following a range of different AED exposures was calculated. The MCM rate was calculated as shown below:

$$\frac{\text{the number of live births with an MCM} + \text{the number of pregnancy losses with an MCM}}{\text{the total number of live births} + \text{the number of pregnancy losses with an MCM}}$$

Pregnancy losses (spontaneous abortions, elective terminations and stillbirths) without an MCM were excluded from the denominator. This approach is commonly taken by pregnancy registries because of the likelihood of inconsistent identification of defects across pregnancy losses.^(14,23) Multiple births were included in the analyses but it was decided *a priori* that if there was a case

of a multiple birth where both/all infants had the same MCM they would only be counted once in the numerator and once in the denominator. Relative risks (RRs) with 95% confidence intervals (CIs) were calculated using STATA (version 9; StataCorp LP, College Station, TX, USA) to compare the risk of a pregnancy outcome with an MCM following mono- and polytherapy AED exposure with those who had no AED exposure during the first trimester. Comparisons of MCM risk were also made for the three most commonly used AEDs. For this, carbamazepine was selected as the baseline prevalence as it was the comparator chosen in the UK register analyses and it had the largest number of exposures. Comparisons with the UK Epilepsy and Pregnancy Register were made with the figures that they reported in their publication of December 2006.⁽¹⁴⁾ The above analyses were also completed to investigate the association between first-trimester monotherapy exposure to valproate and spina bifida in the GPRD.

7.3 Results

7.3.1 Study cohort

The GPRD study cohort is depicted in Figure 7.1. The cohort consisted of 2019 live mother-baby pairs, 551 pregnancy terminations, 13 stillbirths and 1 neonatal death. Approximately 49% of the mother-baby pair cohort had been exposed to ≥ 1 AED during the first trimester of pregnancy. For mother-baby pairs who had at least 2 years of data before the start of pregnancy ($n = 1497$), there was evidence that 15.8% had discontinued AED use in the 2 years before becoming pregnant. Of those exposed to an AED during the first trimester, 83.2% were exposed to monotherapy and 16.8% to polytherapy. The three AEDs most frequently prescribed as monotherapy were carbamazepine, valproate and lamotrigine (Table 7.1).

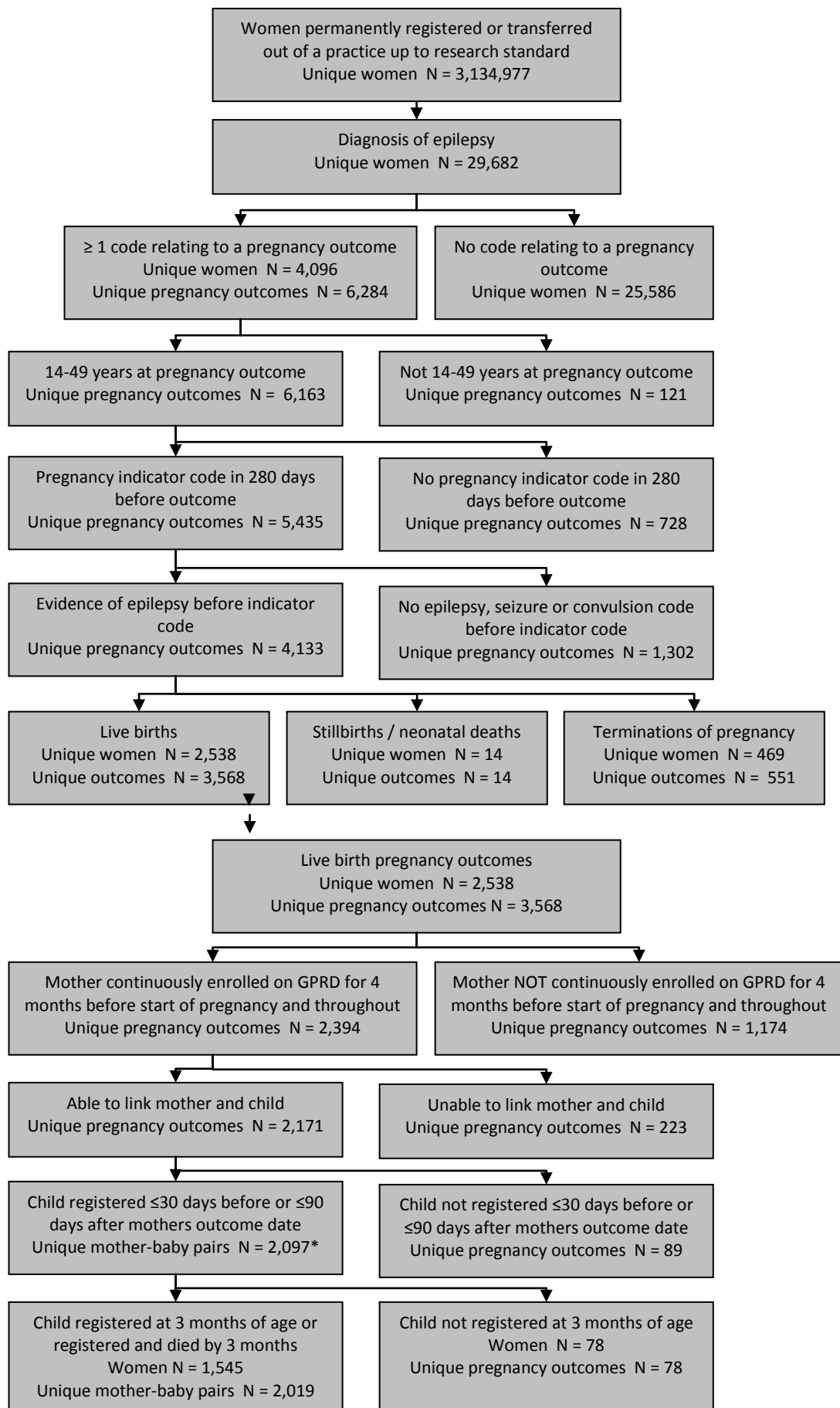


Figure 7.1 Identifying eligible pregnancies and mother-baby pairs * including twins/triplets

Table 7.1 Population characteristics for mother-baby pairs where offspring were registered on the General Practice Research Database

Characteristic	Live births registered at 3 months of age ^a	Live births registered at 1 year of age ^b
No. of mother-offspring pairs	2019	1766
Unique women	1545	1355
Average age at pregnancy outcome [years; mean (SD)]	30.0 (5.6)	30.1 (5.6)
Sex of offspring – male (%)	51.2	51.1
Average number of years infant was followed up/present on the database (SD)	5.4 (4.0)	6.1 (3.9)
First-trimester exposure [n (%)]	994 (49.2)	871 (49.3)
monotherapy	818	706
polytherapy	164	154
drug switching	12	11
Unexposed	1025	895
First-trimester monotherapy		
carbamazepine	355	310
valproate	248	221
lamotrigine	128	98
phenytoin	57	52
phenobarbitone	11	11
clonazepam	6	4
topiramate	5	4
gabapentin	4	2
ethosuximide	3	3
levetiracetam	1	1

^a Or registered and died before reaching 3 months of age.

^b Or registered and died before reaching 1 year of age.

7.3.2 Major congenital malformations

MCMs by pregnancy outcome

A total of 82 MCMs in 62 unique live-born individuals were identified and met the confirmation criteria. Review of the free text confirmed 14 pregnancies met the inclusion criteria and had been terminated following a prenatal MCM diagnosis (Table 7.2). No further MCMs were identified following review of the free text for either the stillbirths or neonatal deaths.

Table 7.2 Number of malformations identified meeting the major congenital malformation inclusion criteria

Malformation	No. of malformations (live births)	No. of malformations (terminations)
Neural tube defect		
Hydrocephalus	2	1
Spina bifida	1	6
Anencephaly	0	2
Cardiac		2 ^a
Atrial septal defect	4	0
Ostium secundum atrial septal defect	2	0
Coarctation of aorta	1	0
Patent ductus arteriosus	1	0
Pulmonary stenosis	3	0
Pulmonary artery atresia	1	0
Tetralogy of Fallot	3	0
Ventricular septal defect	10	0
Transposition of great arteries	0	1
Facial cleft		
Cleft palate	3	0
Cleft lip and palate	2	0
Hypospadias/genitourinary tract		
Absence of kidney	1	0
Atrophy of kidney	1	0
Hydronephrosis	5	0
Hypospadias	11	0
Cystic kidney disease	1	0
Renal tract disorder	0	1
Gastrointestinal tract defects		
Hirschsprung's disease	1	0
Imperforate anus	2	0
Tracheo-oesophageal fistula	1	0
Other		
Congenital cataract	1	0
Foetal valproate syndrome	4	0
Hip dislocation/dysplasia	3	0
Limb reduction	4	0
Talipes equinovarus	4	0
Talipes unspecified	1	0
Polydactyly	3	0
Syndactyly	2	0
Multiple abnormalities	0	1
Reference to 'congenital anomaly', i.e. club hand, club foot, syndactyly and limb defect"? See also Table 7.3	4	0

^a Unable to determine exact malformation type but given that they resulted in an induced pregnancy termination they have been assumed to be major.

By infant follow-up

Of the 82 confirmed MCMs, 53 (64.6%) had been recorded in the GPRD by 3 months of age and 71 (86.6%) by 1 year of age; a breakdown is shown in Table 7.3. Given that <65% of confirmed MCMs had been recorded during the first 3 months of life, it was decided to calculate the prevalence of MCMs at 1 year of age. The analysis cohort therefore included only those infants still present in the GPRD at age 1 year, those who had registered and died by 1 year of age and those pregnancy losses with a non-chromosomal MCM. Of the 2019 mother-baby pairs within our initial cohort, 1766 (87.5%) met these criteria and, including terminations, there were 62 unique pregnancy outcomes with an MCM. The characteristics of mother-baby pairs still present in the GPRD at 1 year of age were not found to differ substantially from those who were still present at 3 months of age (Table 7.1).

7.3.3 Prevalence of major congenital malformations

Type of antiepileptic drug therapy

Table 7.4 shows the prevalence of MCMs diagnosed by 1 year of age following different types of first-trimester AED exposure. An increased risk of an MCM was observed in the GPRD following first-trimester polytherapy exposure when compared with women having no AED exposure. The point estimates (absolute risks) for each exposure category were similar for the GPRD and the UK register; however, in the GPRD, the increased risk following polytherapy compared with monotherapy did not reach statistical significance (RR 1.73; $p = 0.10$ compared with RR 1.63; $p = 0.01$).

Table 7.3 The age by which malformations in live-born infants were recorded in the General Practice Research Database

Malformation	Number of live births with MCMs	Recorded by 3 months	Recorded after 3 months / before 1 year	Recorded after 1 year
Neural tube defect				
Hydrocephalus	2	0	1	1
Spina bifida	1	1	0	0
Cardiac				
Atrial septal defect	4	3	0	1
Ostium secundum atrial septal defect	2	2	0	0
Coarctation of aorta	1	0	1	0
Patent ductus arteriosus	1	1	0	0
Pulmonary stenosis	3	2	0	1
Pulmonary artery atresia	1	1	0	0
Tetralogy of Fallot	3	3	0	0
Ventricular septal defect	10	6	4	0
Facial cleft				
Cleft palate	3	3	0	0
Cleft lip and palate	2	2	0	0
Hypospadias/Genitourinary tract				
Absence of kidney	1	0	0	1
Atrophy of kidney	1	0	1	0
Hydronephrosis	5	5	0	0
Hypospadias	11	8	1	2
Cystic kidney disease	1	0	0	1
Gastrointestinal tract defects				
Hirschsprung's disease	1	0	0	1
Imperforate anus	2	2	0	0
Tracheo-oesophageal fistula	1	1	0	0
Other				
Congenital cataract	1	0	1	0
Foetal valproate syndrome	4	1	1	2
Hip dislocation/dysplasia	3	3	0	0
Limb reduction	4	1	3	0
Talipes equinovarus	4	2	1	1
Talipes unspecified	1	1	0	0
Polydactyly	3	1	2	0
Syndactyly	2	1	1	0
Reference to 'congenital anomaly' i.e. club hand, club foot, syndactyly and limb defect	4	3	1	0

Table 7.4 Prevalence of major congenital malformations (MCMs) following different types of first-trimester antiepileptic drug exposure

General Practice Research Database				UK Epilepsy and Pregnancy Register		
Exposure type	No. of exposures	Unique offspring with MCMs ^a [n (%)]	Relative risk (95% CI)	No. of exposures	Unique offspring with MCMs ^b [n (%)]	Unadjusted odds ratio (95% CI)
Unexposed	902	22 (2.4)	1	227	8 (3.5)	1
Monotherapy	711	29 (4.1)	1.67 (0.97, 2.89); p = 0.060	2468	91 (3.7)	1.05 (0.50, 2.19); p = 0.90
Polytherapy	156	11 (7.1)	2.89 (1.43, 5.84); p = 0.005	718	43 (6.0)	1.71 (0.79, 3.69); p = 0.17

^a Recorded in child's electronic medical record by 1 year of age.

^b Outcome data collected approximately 3 months after expected delivery date.

7.3.4 Monotherapy exposures

Table 7.5 shows the prevalence of MCMs for the three most commonly prescribed monotherapy exposures. The GPRD observed a higher carbamazepine MCM prevalence than was found in the UK register, which, surprisingly, found a lower prevalence of MCMs in pregnancies exposed to carbamazepine than in unexposed pregnancies. As can be seen from Table 7.5, this had implications for the RR estimates of exposure when carbamazepine was used as the baseline prevalence. Bearing in mind the known issues with selection bias to pregnancy registers, we considered the possibility that for the pregnancy register, carbamazepine may have been an inappropriate reference group and it might have been more appropriate to compare all exposures with the unexposed population. Owing to the lower carbamazepine MCM prevalence, we therefore calculated the RR estimates both for the GPRD and the UK register, with the unexposed populations as the baseline prevalence (Table 7.5). As a result, risk estimates for carbamazepine and lamotrigine were in the opposite direction in the GPRD compared with the UK register, and the point estimate for valproate was higher in the GPRD than in the UK register, although, as in the register, it did not reach statistical significance (RR 2.00; 95% CI 0.99, 4.07; $p = 0.05$).

7.3.5 Valproate and spina bifida

The UK register did not report on the rate of spina bifida among monotherapy valproate-exposed pregnancies. It did, however, publish the rate of all neural tube defects following monotherapy valproate exposures, and this was estimated at 1.0% compared with 2.2% in the GPRD. Further analysis in the GPRD identified seven cases of spina bifida; one was a live birth and six were terminations of pregnancy. Of these, during the first trimester, four cases were exposed to valproate in monotherapy, one was exposed to other AEDs in polytherapy and two were unexposed to any AEDs. The prevalence of spina bifida in valproate monotherapy-exposed pregnancies (resulting in a live birth, stillbirth or termination of pregnancy) was 1.78% (95% CI 0.05, 3.50) compared with 0.22% (95% CI 0.00, 0.53) for pregnancies with no first-trimester AED exposure. The prevalence of spina bifida was significantly higher in the valproate-

Table 7.5 Prevalence of major congenital malformations (MCMs) following different monotherapy first-trimester antiepileptic drug exposures with carbamazepine as the reference category and then with the unexposed as the reference category

General Practice Research Database				UK Epilepsy and Pregnancy Register		
Monotherapy exposure	Number of exposures (%)	Unique offspring with MCMs ^a [n (%)]	Relative risk (95% CI)	Number of exposures (%)	Unique offspring with MCMs ^b [n (%)]	Unadjusted odds ratio (95% CI)
Carbamazepine	311 (49.1)	13 (4.2)	Reference	900 (39.8)	20 (2.2)	Reference
Lamotrigine	98 (15.5)	3 (3.1)	0.73 (0.21, 2.52); p = 0.77	647 (28.6)	21 (3.2)	1.44 (0.77, 2.67); p = 0.25
Valproate	225 (35.5)	11 (4.9)	1.17 (0.53, 2.56); p = 0.70	715 (31.6)	44 (6.2)	2.78 (1.62, 4.76); p<0.001
Unexposed	902	22 (2.4)	Reference	227	8 (3.5)	Reference ^c
Carbamazepine	311	13 (4.2)	1.71 (0.87, 3.36); p = 0.11	900	20 (2.2)	0.63 (0.28, 1.41); p = 0.26
Lamotrigine	98	3 (3.1)	1.26 (0.38, 4.17); p = 0.73	647	21 (3.2)	0.92 (0.41, 2.05 ; p = 0.84
Valproate	225	11 (4.9)	2.00 (0.99, 4.07); p = 0.05	715	44 (6.2)	1.75 (0.83, 3.65); p = 0.13

^a Recorded in child's electronic medical record by 1 year of age.

^b Outcome data collected approximately 3 months after expected delivery date.

^c Calculated by the authors and not reported by the UK Epilepsy and Pregnancy Register.

exposed group compared with those unexposed to AEDs (RR 8.02; 95% CI 1.5, 43.5).

7.4 Discussion

In this study we were able to identify a known teratogenic association in the GPRD, suggesting there may be instances where the GPRD can be used for the evaluation of drug safety in pregnancy. Differences were found, however, when comparing overall MCM rates for AEDs in the GPRD with UK register data. Below we discuss some of the key findings along with strengths and limitations of using the GPRD as a tool for postmarketing teratogen surveillance.

7.4.1 Sample size

The relatively small number of AED exposures in the GPRD was a limitation that resulted in wide confidence intervals around many of the MCM prevalence estimates. It was known from the outset that the number of first-trimester AED exposures in the GPRD was likely to be small, given the low prevalence of epilepsy and the fact that the database captures only about 7% of the UK population. The GPRD is therefore likely to be better suited to more prevalent conditions and the identification of exposures that result in a substantial increase in risk. However, the AED drug class was selected for this study in view of the large amount of comparator information already available from AED pregnancy registries and, more specifically, the presence of the UK register.

7.4.2 Cohort identification and disease status

Not all women with epilepsy will be recorded explicitly as such. Instead, some women with epilepsy will only have records of seizures or convulsions in the GPRD. This results from the possibility that an epilepsy diagnosis may have been made in a hospital setting and not actually recorded within the woman's electronic medical record. GPs are not required to, and do not, record the indication for prescribing with repeat prescriptions. For the purposes of this study, we wanted to ensure we captured all epilepsy cases but excluded women without epilepsy and women who received AEDs for indications other than

epilepsy (e.g. trigeminal neuralgia). Therefore, for women to be taken as having epilepsy, they were required to have either ≥ 2 epilepsy diagnosis codes (without the need for supporting evidence) or a single epilepsy code or ≥ 2 seizure or convulsion codes (plus additional supporting evidence within their medical record).

It is possible that these inclusion criteria may still have resulted in the inclusion of a small number of women in the GPRD cohort who did not actually have epilepsy and therefore would not have been eligible for inclusion in the UK Epilepsy and Pregnancy Register. These women are likely to fall into the 'AED unexposed' category and therefore this may, in part, explain the different proportions in the exposure categories between the two data sources. An alternative or additional explanation for some of the difference in the proportion of exposed versus unexposed pregnancies could be that women with epilepsy who are not taking AEDs, or who have discontinued using AEDs in preparation for their pregnancy, are less likely to choose to enroll and be captured in a pregnancy register. Within the GPRD it was found that 15.8% of those who had ever received an AED prescription, had discontinued AED therapy during the 2 years before becoming pregnant. A number of sensitivity analyses were carried out to assess the effect of different epilepsy inclusion criteria within the GPRD cohort but these were found not to materially alter the risk estimates.

7.4.3 Outcome assessment

It was possible to identify and verify a wide range of congenital malformations in the GPRD. The ability to identify pregnancies in the GPRD with an MCM that resulted in a termination of pregnancy was critical in the identification of the teratogenic association between first-trimester exposure to valproate and an increased risk of a pregnancy outcome with spina bifida. The severe nature of spina bifida and the fact it is often diagnosed prenatally meant that without the terminated pregnancies and supporting free text this association would not have been identified. The inclusion of pregnancy terminations is important for the identification of any severe or life-threatening malformation diagnosed

prenatally. Registries have the benefit of a review of MCM cases by an experienced teratologist, but a potential strength of using the GPRD is the ability to capture more completely than registries those pregnancies that do not result in a live birth.

7.4.4 Exposure assessment

The GPRD has the advantage that prescription data is recorded prospectively and independently of the pregnancy outcome. The level of certainty with which it is possible to determine timing of exposure in the GPRD is limited by the level of precision with which it is possible to estimate the exact LMP date and subsequently the first trimester. The GPRD is therefore more appropriate for medicines used to treat chronic, rather than episodic, conditions where exposure is likely to be continuous. Exposure assessment in the GPRD is, however, purely based on the issue of prescriptions and there is no way of establishing if and when the medicine has been taken. Exposure information in the GPRD is also limited to medicines available for use in the UK that require a prescription; it does not capture over-the-counter medicines.

Both the GPRD and the UK register reported carbamazepine, valproate and lamotrigine as the most commonly used AED monotherapies during the first trimester of pregnancy. The UK register observed a higher proportional representation of lamotrigine exposures than the GPRD. Given that lamotrigine was not launched in the UK until 1992 and would have taken time to penetrate the market, this difference could be the result of the different time periods of data collection, with the GPRD study including pregnancy outcomes from 1990 and the UK register starting in December 1996. Further analyses restricting the GPRD data to the same time period as the UK register, however, still found the register to have a higher proportion of lamotrigine use. This difference could therefore be an example of selection bias within the UK register, with women and healthcare professionals being more likely to report lamotrigine exposures as it is a relatively new drug on the market.

7.4.5 Differences in risk assessment

The UK register reported comparative risk analyses between AED monotherapy-exposure groups using carbamazepine as the reference exposure category. However, the prevalence of MCMs associated with carbamazepine use in the UK register was lower than that observed for the unexposed group and lower than corresponding risk estimates in the GPRD. This, in great part, underlies the differences in comparative results between the UK Register and the GPRD, and illustrates the problem of choosing an appropriate comparator group for these studies of drug safety in pregnancy. Indeed, it highlights the issue of whether a comparator group should be chosen for pregnancy registry studies monitoring for a signal of major teratogenicity where the effects, if present, are likely to be so great they usually will need no comparison groups or consideration of confounding variables.^(24,25)

Irrespective of any comparator group, both data sources reported a lower prevalence of MCMs following first-trimester exposure to valproate than has been reported in a number of studies outside the UK.⁽²⁶⁻³⁰⁾ The reason for this is unclear and further investigation would be required to determine whether it is related to differences in recording, differences in the use of valproate in the UK, differences in the use of concomitant folic acid, or other factors.

7.4.6 Follow-up

Outcome information is often requested by pregnancy registries close to the expected date of delivery (often within the first 3 months of life). Within the GPRD, fewer than 65% of MCMs were recorded by 3 months of age, which may demonstrate a need to follow infants for longer in the GPRD than in registries in order to obtain representative outcome data. Loss to follow-up is an issue with all prospective observational studies. Although in the UK register loss to follow-up was only 8.1%, in some registries it can be as high as 27%.⁽²³⁾ Within the GPRD there is the potential for infant follow-up to extend beyond the period immediately after birth, with the mean length of infant follow-up within this study cohort being 6.1 years (SD 3.9).

7.4.7 Confounders

In order to encourage enrolment, and given that pregnancy registries are designed to identify major teratogens, the amount of information requested in terms of potential confounding factors is often limited. Although not evaluated in this study, information is available within the GPRD on potential confounders such as age, smoking status, alcohol status, body mass index, co-morbidity and other prescription medications; however, further work is required to look at the completeness, accuracy and reliability of these variables.

7.4.8 Comparison of data sources

The UK Epilepsy and Pregnancy Register was set up with an entirely different aim than that of the GPRD. Consequently, the methods of data collection used by computerized medical record databases and pregnancy registries do differ and these differences have implications for the type of information available and the extent to which direct comparisons can be made. Pregnancy registries often rely on voluntary enrolment and therefore may not always capture a truly representative sample of those either with the disease of interest or using the medication of interest. Whilst such selection bias is an issue, the indication for prescribing is usually unambiguous and typically pregnancy registries have detailed information regarding birth defects identified at birth. In contrast, whilst in the GPRD selection bias is negligible or absent, the indication for prescribing is implicit rather than explicit because it is influenced from the presence of diagnostic and symptom records and the absence of alternative explanations. In addition, to obtain detailed information regarding malformations identified, exploring the electronic records does not suffice. Instead, free text or hospital letters need to be consulted.⁽²⁰⁾

7.5 Conclusions

Postmarketing surveillance of pregnancy outcomes to identify potential teratogens is essential because of the lack of evidence available when a new product is first marketed regarding its safety when used during pregnancy. The GPRD has proven useful in the identification of malformations and of a major

teratogenic association. The GPRD does, however, identify fewer exposed pregnancies than a pregnancy registry, especially for less prevalent conditions; therefore, in many circumstances pregnancy registries are likely to remain the optimum method of surveillance. Given limitations in sample size, the GPRD is going to be most capable of identifying major risk factors rather than those that result in a relatively small increase in risk. The GPRD may also be better suited to monitoring medicines used to treat more prevalent conditions, such as depression, or medicines that have been on the market for a long time for which no registry has been set up.

7.6 References

1. Honein MA, Paulozzi LJ, Cragan JD, et al. Evaluation of selected characteristics of pregnancy drug registries. *Teratology* 1999;60:356-64
2. Schirm E, Meijer WM, Tobi H, et al. Drug use by pregnant women and comparable non-pregnant women in The Netherlands with reference to the Australian classification system. *Eur J Obstet Gynecol Reprod Biol* 2004;114(2):182-8
3. Lakha F, Glasier A. Unintended pregnancy and use of emergency contraception among a large cohort of women attending for antenatal care or abortion in Scotland. *Lancet* 2006;368(9549):1782-7
4. Henshaw SK. Unintended pregnancy in the United States. *Fam Plann Perspect* 1998;30(1):24-9
5. Koren G, Pastuszak A, Ito S. Drugs in pregnancy. *N Engl J Med* 1998;338(16):1128-37
6. Mitchell A. Special considerations in studies of drug-induced birth defects. In: Strom BL, editor. *Pharmacoepidemiology*. 3rd ed. Chichester: John Wiley and Sons Ltd
7. Charlton RA, Cunningham MC, de Vries CS, et al. Data resources for investigating drug exposure during pregnancy and associated outcomes: the General Practice Research Database (GPRD) as an alternative to pregnancy registries. *Drug Saf* 2008;31(1):39-51
8. Hardy JR, Holford TR, Hall GC, et al. Strategies for identifying pregnancies in the automated medical records of the General Practice Research Database. *Pharmacoepidemiol Drug Saf* 2004;13:749-59
9. Snowball JM, de Vries CS. Determination of pregnancy on the General Practice Research Database. *Pharmacoepidemiol Drug Saf* 2007;16:S118
10. Devine S, West S, Andrews E, et al. The identification of pregnancies within the General Practice Research Database. *Pharmacoepidemiol Drug Saf* 2010 19(1):45-50

11. Hardy JR, Leaderer BP, Holford TR, et al. Safety of medications prescribed before and during early pregnancy in a cohort of 81 975 mothers from the UK General Practice Research Database. *Pharmacoepidemiol Drug Saf* 2006;15:555-64
12. Wurst KE, Ephross SA, Loehr J, et al. The utility of the general practice research database to examine selected congenital heart defects: a validation study. *Pharmacoepidemiology Drug Saf* 2007;16:867-77
13. Devine S, West SL, Andrews E, et al. Validation of neural tube defects in the full featured-general practice research database. *Pharmacoepidemiol Drug Saf* 2008;17:434-44
14. Morrow J, Russell A, Guthrie E, et al. Malformation risks of antiepileptic drugs in pregnancy: a prospective study from the UK Epilepsy and Pregnancy Register. *J Neurol Neurosurg Psychiatry* 2006;77(2):193-8
15. Robert E, Guibaud P. Maternal valproic acid and congenital neural tube defects [letter]. *Lancet* 1982; II: 937
16. Lindhout D, Meinardi H. Spina bifida and in-utero exposure to valproate. *Lancet* 1984; II: 396
17. Dolk H, Jentink J, Loane M, et al., on behalf of the EUROCAT Antiepileptic Drug Working Group. Does lamotrigine use in pregnancy increase orofacial cleft risk relative to other malformations? *Neurology* 2008;71:714-22
18. Wood L, Martinez C. The General Practice Research Database: role in pharmacovigilance. *Drug Saf* 2004;27(12):871-81
19. MHRA. The General Practice Research Database (2009) [online]. Available from URL: <http://www.gprd.com/home/default.asp> [Accessed 01.03.2010]
20. Charlton RA, Weil JG, Cunningham MC, et al. Identifying major congenital malformations in the UK General Practice Research Database (GPRD): a study reporting on the sensitivity and added value of photocopied medical records and free text in the GPRD. *Drug Saf* 2010;33(9):741-50
21. EUROCAT. List of minor anomalies for exclusion up to birth year 2004 (2000) [online]. Available from URL: <http://www.eurocat-network.eu/content/EUROCAT-Old-List-Minor-Anomalies.pdf> [Accessed 06.02.2010]
22. EUROCAT. Minor anomalies for exclusion (chapter 3.2, guide 1.3) 2005 [online]. Available from URL: <http://www.eurocat-network.eu/content/EUROCAT-Guide-1.3.pdf> [Accessed 06.02.2010]
23. Kendle International Inc. The lamotrigine pregnancy registry: interim report, 1 September 1992 through 31 March 2009 [online]. Available from URL: http://pregnancyregistry.gsk.com/documents/lam_report_spring2009.pdf [Accessed 06.02.2010]
24. Mitchell A. Systematic identification of drugs that cause birth defects: a new opportunity. *N Engl J Med* 2003;349(26):2556-9
25. Charlton RA, Cunningham M, Weil JG, et al. Regarding the need, or lack thereof, of a comparator group for pregnancy registries [letter]. *Birth Defects Res A Clin Mol Teratol* 2009;85 808
26. Samren EB, van Duijn CM, Christiaens GCML, et al. Antiepileptic drug regimens and major congenital abnormalities in the offspring. *Ann Neurol* 1999;46:739-46
27. Kaneko S, Battino D, Andermann E, et al. Congenital malformations due to antiepileptic drugs. *Epilepsy Res* 1999;33(2-3):145-58
28. Wide K, Winbladh B, Kallen B. Major malformations in infants exposed to antiepileptic drugs in utero, with emphasis on carbamazepine and valproic acid: a nation-wide, population-based register study. *Acta Paediatr* 2004;93:174-6

29. Wyszynski DF, Nambisan M, Surve T, et al. Increased rate of major malformations in offspring exposed to valproate during pregnancy. *Neurology* 2005;64(6):961-5
30. Vajda FJE, O'Brien T, Hitchcock A, et al. Antiepileptic drugs in pregnancy: great promise from pregnancy registries. *J Clin Neurosci* 2005;10:543-9
31. Hart AC. Hopkins CA. editors. ICD-9-CM professional for hospitals. Volumes 1,2 and 3. International classification of diseases 9th revision, clinical modification 6th ed. Reston (VA): Ingenix Inc. 2003. St Anthony Publishing/Medicode, 2003

Chapter 8

Recording of pregnancy losses on the General Practice Research Database

8.1 Introduction

Post marketing surveillance for the detection of potential teratogens and other harmful effects is essential given the limited amount of information available when a new medicine is introduced on the market. It has been reported that approximately 37% of clinically recognised pregnancies end in a spontaneous pregnancy loss or an induced termination,⁽¹⁾ emphasising the need for these types of pregnancy outcome to be monitored in addition to live- and still- births when evaluating the safety of a medicine used during pregnancy. Over the past two decades there have been major advances in prenatal screening^(2, 3) and now it is increasingly likely that some severe congenital malformations, such as neural tube defects, will be diagnosed prenatally. Women may subsequently opt to have their pregnancy terminated. The fact that teratogens are likely to increase the risk of selected defects, rather than all defects,⁽⁴⁾ means it is possible that if the selected defect type can be diagnosed via prenatal screening the increase in risk may go undetected, if terminations of pregnancy are not captured in postmarketing safety evaluations. The optimum data sources for post marketing teratogen surveillance will therefore be those that can capture all types of pregnancy outcome and distinguish between induced pregnancy terminations that were carried out for medical and non-medical reasons.

The General Practice Research Database (GPRD) has been identified as a potential source for monitoring pregnancies and pregnancy outcomes.⁽⁵⁻⁸⁾ The importance of including terminations of pregnancy when evaluating pregnancy outcomes in the GPRD has been demonstrated for neural tube defects in general⁽⁹⁾ and in relation to first trimester exposure to valproate and a pregnancy outcome with spina bifida specifically.⁽¹⁰⁾ Hardy *et al.*⁽⁸⁾ and Devine *et al.*⁽⁷⁾ have both reported on the generation of an algorithm to identify different types of pregnancy outcome within the GPRD. In this chapter we focus on pregnancy loss and describe the development of an algorithm to categorise pregnancy losses into those that were spontaneous and those that were induced and for those that were induced, whether or not they were carried out following prenatal diagnosis of a congenital malformation (including chromosomal abnormalities).

8.2 Methods

8.2.1 Data source

The GPRD is the world's largest computerised database of anonymised, longitudinal medical records from primary care.⁽¹¹⁾ At present it contains over 57 million person years of data and is actively collecting data on approximately 4 million patients (~7% of the UK population) registered with around 500 GP practices within the UK.⁽¹²⁾ The GPRD consists largely of coded data entered onto a computer system by GPs as part of the clinical management of patients. There are over 3,500 different medical codes relating to pregnancy that a GP can enter into a patient's record on the database. In addition to pregnancy related data the GPRD also captures other medical diagnoses, prescriptions, referrals and consultations. As well as recording medical codes, GPs have the option of recording additional, non-coded, information relating to a particular diagnosis in the free text field. This free text is not readily available to researchers but can be purchased from the database provider separately on a per project basis.

8.2.2 Identification of pregnancy related medical codes

All medical codes relating to pregnancy, childbirth and termination, including those referring to antenatal, neonatal and postnatal care, were identified using a list of search terms and then reviewing those in the hierarchical vicinity. Codes indicating a negative pregnancy test were excluded. Each code was assigned to one of the following categories: delivery (including stillbirths), pregnancy loss (including both spontaneous and induced abortions), a postpartum event, a preterm delivery, a post-term delivery and any remaining codes that provided additional supporting evidence of a pregnancy. Codes were then grouped according to whether they provided sufficient evidence of a pregnancy or whether additional supporting evidence in the form of another pregnancy related code was required.

8.2.3 Study population

The study period ran from 1 January 1992 until 18 February 2009. The study population consisted of females with ≥ 1 code indicating a pregnancy potentially ending between January 1992 and their right censor date (date of last data collection, date the patient died or left the practice or the date the practice stopped contributing data considered by the GPRD group at the Medicines and Healthcare products Regulatory Agency to be up-to-standard (UTS) for the purposes of research). Pregnancies occurring before a patient's left censor date (latest of registration date or date the practice was first considered to be providing data that is UTS) were excluded.

8.2.4 Identification and classification of pregnancy losses

All pregnancy outcomes were identified in the GPRD and categorised as being a delivery (including stillbirths) or a pregnancy loss. For each patient with a medical code that required supporting evidence, a check was made to ensure at least one additional, different pregnancy related code was present within a specified time frame (which was dependent on the category of the code).

For each patient, pregnancy loss events were assessed in ascending date order with the last event in a series of 6 weeks being taken as the pregnancy end date. This was because earlier codes of pregnancy loss commonly related to requests and referrals for a termination rather than the date of the termination itself. As a consequence, where two pregnancy loss events occurred in short succession but with more than six weeks between the first and last recording, the second pregnancy loss was omitted from our analyses.

The start date of the pregnancy was determined, where possible, from the following records in order of priority; estimated date of delivery (EDD), EDD calculated from record for last menstrual period (LMP), EDD calculated from delayed/missed period, default pregnancy length for a pregnancy loss (10 weeks). There were a number of apparent deliveries (codes included 'delivery details' and 'birth details') that were contradicted by the existence of miscarriage

records within days of the coded 'delivery'. Manual review of a sample of these patients' medical records made it apparent there were instances where GPs recorded codes implying a delivery when actually it was a pregnancy loss. Where a record for a delivery had a code for a spontaneous or induced pregnancy loss within the 42 days before or 21 days after, the delivery record was disregarded.

An algorithm was created to categorise, with varying levels of certainty, a pregnancy loss relating to either an ectopic pregnancy, a hydatidiform mole, a spontaneous abortion or an induced termination. Where we were unable to classify pregnancy losses, the type of loss was classed as unknown. For losses categorised as being induced terminations, medical codes and prescriptions were scrutinised to determine whether the termination was induced following a prenatal diagnosis of a congenital malformation or for non-medical reasons.

8.2.5 Review of free text comments

To evaluate the reliability of the algorithm, free text comments recorded during the two months before a pregnancy loss and four months after were requested and reviewed for a sample of 200 pregnancy losses.

8.2.6 Data analysis

The percentage of identified pregnancies that resulted in a delivery and a pregnancy loss was calculated. The number and percentage of pregnancy losses that were identified at each stage of the algorithm were then grouped into the different types of pregnancy loss. The percentage of identified induced terminations with evidence to suggest they were carried out for a foetal medical condition was also calculated and compared to national statistics. To increase the likelihood that only true pregnancies were included, those pregnancies where the female was not 11-50 years of age at the time of the pregnancy outcome were excluded from the analyses. Comparisons were made between the algorithm end point of a pregnancy loss and the information recorded in the free text comments.

8.3 Results

8.3.1 Identification of pregnancy losses

In total 529,975 pregnancies were identified. Of these, 370,528 (69.9%) resulted in a delivery and 159,447 (30.1%) in a pregnancy loss. For deliveries and pregnancy losses the mean age at pregnancy outcome was 29.8 years ($SD^{++} = 5.8$) and 28.8 years ($SD = 7.4$) respectively.

Figure 8.1 shows the number of pregnancy losses identified at each stage of the algorithm. The number and percentage of pregnancy losses that could be categorised as being spontaneous, induced and unknown are shown in Table 8.1.

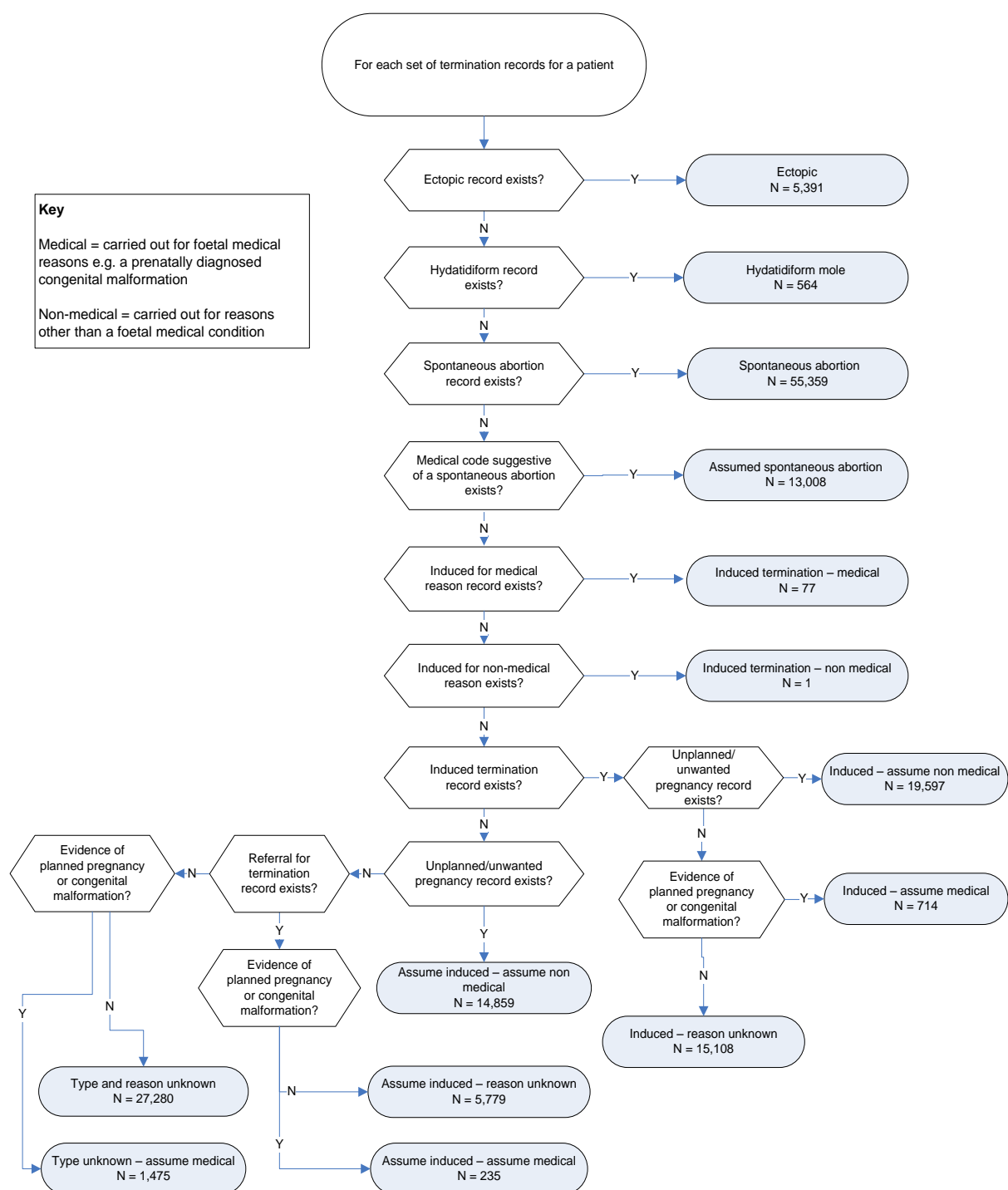
8.3.2 Categorisation of pregnancy losses

For 82.0% of all pregnancy losses sufficient information was recorded to categorise the pregnancy loss as being either spontaneous, ectopic, a hydatidiform mole pregnancy or induced. For the remaining 18.0% there was insufficient evidence to determine the type of pregnancy loss.

For pregnancy losses recorded as being induced (or assumed induced) terminations it was found that medical codes specifically stating the reason for the termination were recorded in less than 0.2% of cases. In 62.8% of cases, however, there were medical codes providing evidence of a planned/wanted pregnancy, an unplanned/unwanted pregnancy or a congenital malformation, enabling an assumption to be made as to the reason for the termination.

Of all pregnancy losses categorised as being induced (or assumed induced), 0.1% had medical codes specifically stating they were carried out owing to a medical indication. In addition, a further 1.7% had medical codes relating to the presence of a congenital malformation or providing evidence that it was either a planned/wanted pregnancy, allowing the assumption to be made that the pregnancy was terminated for medical reasons.

⁺⁺ Standard deviation



Examples of medical codes

Spontaneous abortion e.g. 'miscarriage', 'spontaneous abortion', 'complete spontaneous abortion'
 Medical code suggestive of spontaneous abortion e.g. 'missed abortion', 'retained products of conception removed'
 Induced for medical reasons e.g. 'termination of pregnancy – medical indication', 'therapeutic abortion'
 Induced for non-medical reasons e.g. 'abortion induced – social reasons'
 Induced termination e.g. 'legally induced abortion',
 Evidence of unplanned or unwanted pregnancy e.g. 'pregnant – unplanned', 'unwanted pregnancy'
 Evidence of planned pregnancy e.g. 'planned pregnancy', 'prescriptions for pre-conceptional folic acid, treatment for infertility or prevention of miscarriage'
 Congenital malformation e.g. a code for a congenital malformation recorded during pregnancy or in the 30 days after the termination date

Figure 8.1 The number of pregnancy losses identified at each endpoint of the algorithm

Table 8.1 Breakdown of the number of different types of pregnancy loss identified

Type of pregnancy loss	Number of pregnancy losses	Percentage of pregnancy losses N = 159,447	Percentage of all pregnancies N = 529,975
Spontaneous	68,367	42.8	12.9
Of which -			
- specific spontaneous / miscarriage code	55,359	34.6	
- contributory evidence was spontaneous ¹	13,008	8.1	
Ectopic	5,391	3.4	1.0
Hydatidiform mole	564	0.4	0.1
Induced / assumed induced	56,370	35.4	10.6
- for medical reasons ²	77	<0.1	
- assume for medical reasons ³	949	0.6	
- for non-medical reasons ⁴	1	<0.1	
- assume for non-medical reasons ⁵	34,456	21.6	
- reason unknown	20,887	13.2	
Unknown	28,755	18.0	5.4
- assume not induced for non-medical reasons ⁶	1,475	0.9	
- reason unknown	27,280	17.1	

¹ Codes including 'missed abortion', 'retained products of conception removed'

² Codes including 'termination of pregnancy – medical indication', 'therapeutic abortion'

³ Based on evidence of a planned or wanted pregnancy, pre conceptional folic acid prescriptions, a congenital malformation, amniocentesis, treatment for infertility

⁴ Codes including 'abortion induced – social reasons'

⁵ Based on evidence of an unwanted or unplanned pregnancy e.g. 'pregnancy – unwanted'

⁶ Evidence of a planned pregnancy; type of loss therefore assumed either spontaneous or induced for medical reasons

8.3.3 Verification of the algorithm

Review of the free text comments for 200 pregnancy losses found for 108 pregnancies the free text provided confirmation of the algorithm endpoint. For a further 51 pregnancy losses that the algorithm had categorised as type and/or reason unknown, the free text provided additional information enabling the type and reason for the pregnancy loss to be determined. Four of these were found to be induced for medical reasons. In 7 pregnancies the free text contained information that contradicted the type of pregnancy loss derived by the algorithm and for the remaining 34 pregnancy losses the free text contained no relevant additional information (Table 8.2).

Table 8.2 Comparison of algorithm end points with those derived from review of the free text comments

Type of pregnancy loss based on the algorithm categorisation	Type of pregnancy loss based on review of free text comments					
		Spontaneous	Induced for medical reasons	Induced for non-medical reasons	Induced – reason unknown	Type and reason unknown
	Spontaneous	19		2	2	4
	Induced for medical reasons		3	2	1	1
	Induced for non-medical reasons		1	64	12	10
	Induced – reason unknown		1	20	13	6
	Type and reason unknown		3	19	8	9

Key

- Algorithm verified using free text
- Free text contradicts algorithm
- Free text provides additional information
- No further additional insights gained from free text

Summary of algorithm end points

Spontaneous = Ectopic + Hydatidiform mole + Spontaneous abortion + Assumed spontaneous abortion

Induced for medical reasons = Induced-medical + Induced assume medical + Assumed induced assume medical + Unknown assume medical

Induced for non-medical reasons = Induced-non medical + Induced assume non-medical + Assumed induced assume non-medical + Assume induced non-medical

Induced – reason unknown = Induced unknown + Assume induced unknown

8.4 Discussion

In this study we developed an algorithm to identify and classify, with varying levels of certainty, pregnancy losses recorded in the GPRD. For 82.0% of pregnancy losses there was a sufficient level of detail available to categorise them as being either spontaneous or induced, whilst for 18.0% the medical codes did not provide sufficient information regularly on the type of pregnancy loss. The percentage of pregnancies falling into each of the pregnancy loss categories in this study corresponded closely with what has been reported in the literature. Of all pregnancies, 12.9% resulted in a spontaneous abortion compared with estimates of between 10 and 20% in the literature.^(13, 14) It is thought the percentage of spontaneous pregnancy losses in the GPRD will be slightly underestimated as a proportion of the pregnancies that the algorithm was unable to classify are likely to have been spontaneous abortions. In addition some early pregnancy losses may never come to the attention of the GP. The algorithm classified 1.0% of all pregnancies in the GPRD as being ectopic and 0.1% as resulting in a hydatidiform mole, both of which correspond with what has been reported elsewhere.⁽¹⁵⁾

It was found that 10.6% of all pregnancies could be categorised as being induced (or assumed induced) terminations. Again this percentage will be slightly underestimated owing to the proportion of pregnancies we were unable to classify with the algorithm developed. Of those classified as induced or assumed induced, 0.1% had a medical code specifically stating that the pregnancy was terminated for a congenital malformation whilst a further 1.7% had medical codes relating to a congenital malformation or the fact the pregnancy was planned and/or wanted, implying it was likely to have been terminated for medical reasons. In England and Wales in 2010, 1.2% of all induced terminations were justified under ground E, "that there was substantial risk that if the child were born it would suffer from such physical or mental abnormalities as to be seriously handicapped."⁽¹⁶⁾ It therefore appears that the algorithm may slightly overestimate the proportion of induced terminations carried out following a diagnosis of a congenital malformation. Alternatively it is possible that not all

terminations carried out due to a congenital malformation are recorded under ground E. In 2010, 97.7% of all terminations were justified under ground C, “that the continuance of the pregnancy would involve risk, greater than if the pregnancy were terminated, of injury to the physical or mental health of the pregnant woman.” It is therefore possible some additional terminations owing to a congenital malformation may have been recorded in this category. In drug safety in pregnancy studies it is critical that all pregnancy outcomes identified as having a congenital malformation are true cases of a congenital malformation. This study and previous research using the GPRD⁽¹⁰⁾ has demonstrated the added value of obtaining free text in order to gain additional information relating to a termination of pregnancy and we recommend requesting this free text for those terminations of pregnancy that have been identified as potentially being induced for a foetal medical condition in order to find further evidence of the prenatal diagnosis.

Unlike Devine *et al.*,⁽⁷⁾ who categorised all pregnancy losses, we were unable to determine the type of pregnancy loss for 18.0% of pregnancy losses identified. Classification of pregnancy losses in the GPRD into those that are spontaneous abortions and those that are induced terminations is complicated by the multiple and sometimes unspecific medical codes that can be used to record the event in the database. In addition pregnancy loss terminology, including the terms ‘abortion’ and ‘termination’ is used interchangeably to describe both a spontaneous and an induced pregnancy loss and this is likely to vary between GPs. It is therefore not possible to determine whether the medical code ‘unspecified abortion’ is referring to a spontaneous or induced pregnancy loss. Identification of supporting evidence in relation to a referral for a termination or a planned pregnancy is therefore important.

A small proportion of clinically recognised pregnancy losses will not be captured in the GPRD. This will consist of some early spontaneous abortions and a small number of induced terminations that were not recorded in the database by the patient’s GP. In 2010, 96% of induced terminations in England and Wales were

funded by the NHS⁽¹⁶⁾ and in the majority of cases the patient will have been referred by their GP so the number that will have been missed is expected to be small.

Owing to the nature of the data source and incomplete data recording within the GPRD assumptions needed to be incorporated into the algorithm and it is possible that in some cases these may have resulted in misclassification of pregnancy outcomes. Initially folic acid prescriptions in the 90 days before and 56 days after LMP were taken as evidence of a planned pregnancy, however on review there were a number of cases where folic acid was prescribed on the same date a woman was requesting a pregnancy termination. An assumption was made that these women were being advised to take folic acid in case they changed their mind about having a termination. The final algorithm was subsequently adjusted to ignore folic acid codes unless it had been prescribed before the first pregnancy code (excluding LMP) was recorded. There may also be examples where a pregnancy was planned, but for a non-medical reason, the woman later changed her mind and decided to opt for a pregnancy termination. In the GPRD, only the planned nature of the pregnancy might have been recorded as a medical code and as a result this pregnancy will have been incorrectly classified as 'assumed induced for assumed medical reasons'.

The extent to which the free text provided supporting evidence was reassuring. Informative free text was available for 166 of the 200 pregnancy losses verified and the type of pregnancy loss, determined by the algorithm, was only contradicted by the free text in 7 cases. The free text provided additional information enabling the categorisation of 51 pregnancy losses where the algorithm had categorised them as type and/or reason unknown and this included 4 that were induced for medical reasons. The cost of free text, however, means it will often not be feasible to request it for all unclassified pregnancy outcomes. For specific drug safety in pregnancy studies however, an algorithm could be developed to identify free text needing to be reviewed based on a search of key words indicative of a type of pregnancy loss (e.g. spontaneous,

induced, elective, therapeutic) or a congenital malformation (e.g. anomaly, malformation, birth defect, spina bifida, anencephaly, amniocentesis etc.). Whilst the algorithm was found to be useful, it is not perfect and for specific drug safety in pregnancy evaluations should be supplemented with review of the free text comments where possible.

8.5 Conclusion

The inclusion of pregnancy losses, especially those induced following a prenatal diagnosis of a congenital malformation, is important when carrying out postmarketing surveillance for potential teratogens. One of the key strengths of the GPRD, unlike some other electronic healthcare databases, is the fact that, although it is incomplete for early loss, it does capture both spontaneous and induced pregnancy losses in addition to deliveries. This study has shown that not only can pregnancy losses be identified in the GPRD but for over 80% it is possible to categorise the type of pregnancy loss and for induced terminations it is possible to identify those that may have been carried out as the result of a foetal medical condition. The GPRD may therefore complement other data sources and provide additional insight in areas where some data sources and methods fall short.

8.6 References

1. Saraiya M, Berg C, Shulman H, et al. Estimates of the annual number of clinically recognised pregnancies in the United States 1981-1991. *Am J Epidemiol.* 1999;149(11):1025-9.
2. Boyd PA, DeVigan C, Khoshnood B, et al. Survey of prenatal screening policies in Europe for structural malformations and chromosome anomalies, and their impact on detection and termination rates for neural tube defects and Down's syndrome. *BJOG.* 2008;115:689-96.
3. Cragan J, Gilboa S. Including prenatal diagnosis in birth defects monitoring: Experience of the Metropolitan Atlanta Birth Defects Program. *Birth Defects Res A Clin Mol Teratol.* 2009;85:20-9.
4. Mitchell A. Studies of drug-induced birth defects. In: Strom BL, editor. *Pharmacoepidemiology*: John Wiley and Sons Ltd; 2005. p. 501-14.
5. Wurst KE, Ephross SA, Loehr J, et al. The utility of the General Practice Research Database to examine selected congenital heart defects: a validation study. *Pharmacoepidemiol Drug Saf.* 2007;16:867-77.
6. Charlton RA, Cunnington MC, de Vries CS, et al. Data resources for investigating drug exposure during pregnancy and associated outcomes: the General Practice Research Database (GPRD) as an alternative to pregnancy registries. *Drug Saf.* 2008;31(1):39-51.
7. Devine S, West S, Andrews E, et al. The identification of pregnancies within the General Practice Research Database. *Pharmacoepidemiol Drug Saf.* 2010;19(1):45-50.
8. Hardy JR, Holford TR, Hall GC, et al. Strategies for identifying pregnancies in the automated medical records of the General Practice Research Database. *Pharmacoepidemiol Drug Saf.* 2004;13:749-59.
9. Devine S, West SL, Andrews E, et al. Validation of neural tube defects in the full featured - General Practice Research Database. *Pharmacoepidemiol Drug Saf.* 2008;17:434-44.
10. Charlton RA, Weil JG, Cunnington M, et al. Comparing the General Practice Research Database and the UK Epilepsy and Pregnancy Register as tools for postmarketing teratogen surveillance: anticonvulsants and the risk of major congenital malformations. *Drug Saf.* 2011;34(2):157-71.
11. Wood L, Martinez C. The General Practice Research Database: Role in pharmacovigilance. *Drug Saf.* 2004;27(12):871-81.
12. MHRA. 2011 The General Practice Research Database. Available from: <http://www.gprd.com/home/> [Accessed on 11.10.2011].
13. Alberman ED. Spontaneous abortion: epidemiology. In: Stabile S, Grudzinkas G, Chard G, editors. *Spontaneous abortion: diagnosis and treatment 1992*. p. 9-20.
14. Wilcox AJ, Weinberg CR, O'Connor JF, et al. Incidence of early loss of pregnancy. *N Engl J Med.* 1988;319(4):189-94.
15. Mishell D, Stenchever M, Droegemueller W, et al. *Comprehensive gynecology*. Missouri: Mosby-Year Book, Inc; 1997.
16. Department of Health 2011. Abortion statistics, England and Wales: 2010. Available from: http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/documents/digitalasset/dh_127202.pdf [Accessed on 11.10.2011].

Chapter 9

General discussion

9.1 Overall discussion of study results

This thesis aimed to evaluate the utility of the United Kingdom's General Practice Research Database to act as an alternative or complement to pregnancy registries. In this chapter I address each of the eight objectives in turn, briefly describe how they were met and discuss the study findings and their implications. In addition I suggest areas for future research as well as summarising some key ongoing developments. Based on all the evidence presented in this thesis I then draw final conclusions on the potential of the GPRD in the field of drug safety in pregnancy research and the areas where it has been demonstrated it can provide a valuable contribution.

Identification of pregnancies in the GPRD

- The first objective was to identify pregnancies recorded in the GPRD within a cohort of women with epilepsy.

Using the algorithm described in Chapters 4 and 6 it was possible to identify pregnancies within a cohort of women with epilepsy in the GPRD. Pregnancies resulting in a live birth, stillbirth, spontaneous pregnancy loss or induced termination were identified and the proportions of pregnancies ending in a delivery and pregnancy loss were found to be in line with what has been reported in the literature.⁽¹⁻³⁾ As outlined in Chapter 8, live births should not be studied in isolation when carrying out drug safety in pregnancy research and therefore the range of pregnancy outcomes captured in the GPRD is a considerable strength over some other data sources.

The algorithm used to identify pregnancies in this study has not, however, been validated and this is something that could be considered in the future. The paper by Devine *et al.* (2009), reporting on a study to identify pregnancies in the GPRD, included two verification exercises to assess the reliability of the number and type of pregnancy outcomes identified.⁽⁴⁾ These found the reliability of live birth identification to be greater than that of pregnancy losses and preterm, post-term and multiple birth outcomes. For the algorithm used in this thesis, in addition to

verifying the type of pregnancy outcome, it would also be of benefit to evaluate the accuracy of LMP dates estimated by the algorithm. This is important as these play a significant role in determining the timing of exposure.⁽⁵⁾ A verification exercise could involve requesting free text comments for a sample of pregnancies to evaluate whether the precise date of the woman's LMP has been entered by the GP as free text and whether there are text comments to support the estimated date of pregnancy outcome. It is likely the introduction of an 'event date' field in the GPs computer software, in addition to an 'entry date' field, will have gone some way to reduce LMP dates being incorrectly recorded on the date the woman visited the GP rather than the date of the actual event but this is something that could also be investigated by review of the free text.

Determining first trimester drug exposure in the GPRD

- The second objective was to identify, within the cohort of women with a diagnosis of epilepsy, those who were exposed to anticonvulsants during the first trimester of pregnancy.

Prescriptions issued in the four months before the estimated LMP date and throughout pregnancy were identified in the women's medical records. The duration of each prescription was mapped based on quantity and daily dose. This was done to ensure women were classified as exposed if they received a prescription before pregnancy that contained enough tablets to allow exposure to continue during the start of pregnancy. Based on my evaluation of the prescribing patterns, where continuous prescribing was identified for most mothers, I conclude that for anticonvulsants, prescription episodes reflected exposure in practice.

For many medicines, estimating the timing of drug exposure can only be as accurate as the estimated timing of the pregnancy start date. When determining the pregnancy start date, attempts were made to utilise all information relating to gestational age, prematurity, estimated date of delivery and LMP date. The level of detail that went into determining pregnancy start dates and mapping

prescriptions will have helped to minimise, although not ruled out, exposure misclassification within our study cohort. Establishing exposure based on the issue of a prescription avoids recall bias, which has the potential to affect the findings of studies carried out using data sources that rely on maternal self reporting.⁽⁶⁾ It does, however, mean it is not possible to obtain information on non-compliance and whether the drug was actually taken. It is for these reasons that GPRD data may be better suited to evaluating the risk of exposures used to treat chronic conditions where discontinuation of exposure is not advised. Epilepsy is a good example of such a condition.

Once women with first trimester exposure to anticonvulsants had been identified it became apparent that the percentage of women with treated epilepsy was considerably lower than expected. Initially we had identified all women with a code relating to epilepsy, seizure or convulsion in an attempt to capture as many as possible. The rationale behind this was that the epilepsy diagnosis could have been made in a hospital and not specifically recorded as a medical code by the woman's GP. The low percentage of anticonvulsant exposure, however, confirmed that this search strategy was too broad and that the presence of a single code within an individual's medical record should not be taken as a guaranteed diagnosis. The initial cohort was used for the study reported in Chapter 6 to allow the recording of as many malformations as possible to be evaluated. For the risk assessment study in Chapter 7, the inclusion criteria was restricted to ensure the cohort was more in line with those enrolled in the UK Epilepsy and Pregnancy Register. For this, supporting evidence in the form of multiple seizure or epilepsy codes or prescription records was required.

Accurate identification of the population under study is important, particularly in risk assessment studies where it is thought the medical condition itself could be a risk factor for adverse pregnancy outcomes. The requirement of supporting evidence will go some way to reduce misclassification but further verification

studies of coding of specific medical conditions recorded in the GPRD are also needed.

Linking the mother's medical record to the child's record

- The third objective was to link, for live deliveries identified in objective one, the mother's medical record to that of the child.

For approximately 90% of the live deliveries it was possible to link the mother to the child. This was based on them having the same family and GP practice numbers, and the child's year and month of birth being equal to the mother's year and month of delivery. This has implications for statistical power when determining whether or not to use the GPRD for evaluations of drug safety in pregnancy. However, it will only impact on the risk estimates if women who cannot be linked to their child differ systematically in terms of risk factors for adverse pregnancy outcomes to those where a link can be made. To my knowledge no study has compared the patient characteristics of women who can and cannot be linked to their child. For drug utilisation studies all exposed women, regardless of a mother-baby link can be included.

Identification and verification of congenital malformations in the GPRD

- The fourth objective was to identify major congenital malformations in offspring of women identified in objective one. Objective five then set out to verify these malformations using either full photocopied medical records or free text comments.

It was possible to identify a wide range of congenital malformations in the GPRD. For malformations identified within the medical record of the child it was possible to verify 85.1% of cases using either the photocopied medical records or free text. Of those, 78.1% had sufficient evidence to be categorised as major congenital malformations, 15.0% were found to be minor anomalies and for 6.9% there was insufficient information to determine the severity of the malformation. When it came to determining the presence of a congenital

malformation, the photocopied medical records were found to be more informative than the free text. Where there was sufficient information to verify the presence of a malformation, however, the percentages that could be categorised as major, minor or unknown were the same for both verification data sources.

The findings of the study presented in Chapter 6 are reassuring in terms of the reliability of malformation recording in the GPRD and they are in line with the study by Wurst *et al.* that verified congenital heart defects.⁽⁷⁾ In future studies, to minimise the likelihood of misclassification of outcomes, it is still advisable to request additional data to verify the cases identified, particularly for malformation types where the severity of the condition is required to determine whether to categorise as major or minor. Although this study found photocopied medical records more informative for verification than free text comments this may change. In recent years, GPs have started to record more information in the free text fields and it is therefore plausible that the extent of information available and the value of free text comments may be greater in the future. Furthermore, the method used in this study to identify the free text for review could be refined to make it more cost and time efficient. This would involve requesting only text recorded in association with a medical code of interest or during a particular time period, rather than requesting all free text recorded within an individual's entire medical record. All work to verify MCMs will not only improve the robustness of the study but will also add to the body of evidence on the accuracy or otherwise of recording in the GPRD.

It is accepted that not all major congenital malformations will be identified via the electronic medical records, however no attempts have been made to quantify this by evaluating the specificity of congenital malformation recording. This is likely to reflect the large amount of time and cost involved in such a task. In 2007, Wurst *et al.* reported on the rates of congenital heart defects in the GPRD compared with those of UK registries reporting to EUROCAT and the National Congenital Anomaly System.⁽⁸⁾ They found the rates to be higher in the

GPRD than both the other sources, which is promising in terms of the specificity of recording although to my knowledge the completeness of reporting to these two systems is unknown. The selection of a suitable comparator with which to evaluate the completeness of recording is difficult, yet more studies evaluating a range of malformations in this way could provide a valuable contribution to the field.

In addition to major congenital malformations, infants can also be born with minor malformations that have little structural significance. Minor malformations are often only reported to pregnancy registries when they are accompanied by a major malformation or there are multiple minor malformations present that could be indicative of a syndrome. The recording of minor malformations was not evaluated in this study using the GPRD, however given their minor clinical significance it is plausible that recording will be less complete and more sporadic than for major malformations.

Comparison with the UK Epilepsy and Pregnancy Register

- The sixth objective aimed to bring together all the aspects of the GPRD that had been evaluated as part of the first five objectives. This involved carrying out a risk assessment study using GPRD data to compare the prevalence of pregnancy outcomes with a major congenital malformation, following a range of first trimester anticonvulsant exposures, with those reported by the UK Epilepsy and Pregnancy Register.

All the relevant information required to carry out a risk assessment study was available from the GPRD. To enable a comparison to be made with the UK Epilepsy and Pregnancy Register, prenatally diagnosed malformations that resulted in a termination of pregnancy were identified in addition to those in live born infants identified in objective four. At the time of this study there was no automated way to identify induced terminations that followed a prenatal diagnosis so free text comments associated with all pregnancy losses were requested and reviewed. This method was found to be effective and identified

14 additional major congenital malformations. It was, however, costly and time consuming and therefore considered an area where more work was required to increase the efficiency with which these malformations could be identified.

It was possible to calculate the risk of a pregnancy outcome with a major congenital malformation following a range of first trimester anticonvulsant exposures, using data from the GPRD. When comparisons were made with the UK Register, what became apparent was that even though the size of the database is large, the number of pregnancies to women exposed to any particular product was small and much lower than that captured by the pregnancy register. These small sample sizes subsequently led to wide confidence intervals and, hence, considerable uncertainty regarding the risk estimates.

The size of the UK population captured by the GPRD is continuously growing and at present is almost twice as large as it was when this study was carried out. Despite this however, for the time being at least, the GPRD will be more suited to evaluating the risks associated with medicines used to treat more prevalent conditions or products that result in a substantial increase in risk. In addition, there are also many prescription products that have been on the market for some time whose safety during pregnancy has never been evaluated. It is possible that these products, even if the teratogenic risk is low, could have the potential for a considerable public health impact if they were commonly used by pregnant women.⁽⁹⁾ The readily available nature of GPRD data may enable the safety of these products to be investigated without the time and expense that would be encountered by using other data sources.

Some differences were observed between the GPRD and the UK Register and these may in part be explained by the differences in data collection methods. The proportional representation of patients exposed to the different anticonvulsants in monotherapy differed between the two data sources. It is possible that this could be the result of selection bias in reporting to the Register

with patients or healthcare professionals being more likely to enrol women exposed to newer products for which less is known about their safety.

The similarity in the risk estimates for monotherapy versus polytherapy exposure for the two data sources was reassuring. Differences were, however, observed in risk estimates for the most commonly reported monotherapy exposures. The fact that lower rates were observed in the GPRD following valproate exposure yet higher rates were observed following carbamazepine exposure makes it difficult to determine the reason behind the discrepancies as given the lower than baseline risk identified with carbamazepine in the Register, it does not appear to be as simple as under- or incorrect reporting in the GPRD.

The UK Register study findings also highlight the issue of the appropriateness of comparator groups. As mentioned in Chapter 2, selecting appropriate comparator groups when evaluating data from pregnancy exposure registries is difficult. For conditions where discontinuing drug therapy is not an option, it may be more informative to make comparisons with other treatment groups as was done by the UK Register. This is valid as long as women with more serious epilepsy who require more continuous monitoring by epilepsy specialists are not more likely to enrol than those whose epilepsy is well controlled. An advantage of the GPRD is the ability to have multiple comparator groups that have been recruited in the same way without adding much additional time and cost to a study. In addition, the representativeness of the population captured by the GPRD is a great strength and in many cases will ensure a greater generalisability of study findings to the wider population.

In addition to unadjusted risk estimates, the UK Register also reported risk estimates adjusted for age at delivery, parity, family history of a congenital malformation, periconceptional folic acid and the sex of the infant. The work presented in Chapter 7 did not attempt to calculate adjusted risk estimates for the GPRD data. Data sources that involve a maternal interview have the benefit of being able to obtain accurate information on factors such as parity, a family

history of congenital malformations and folic acid exposure, which in some cases may have been purchased over the counter.

Within the GPRD data is available on some potential risk factors. These include alcohol intake, smoking status, body mass index, socioeconomic status, co-prescribing and co-morbidities. To my knowledge, only one study has evaluated alcohol and smoking recording in the GPRD.⁽¹⁰⁾ The percentage of individuals who could be identified as a smoker in the GPRD was lower than that reported by the UK Office for National Statistics, whilst the percentage classified as an alcohol drinker was similar for the two sources. More work, however, is required in verifying the accuracy of the recording of this information and the algorithms used to categorise individuals.

Identification of a known teratogenic association in the GPRD

- The seventh objective was to determine whether using GPRD data it was possible to identify the known teratogenic association between monotherapy first trimester exposure to valproate and an increased risk of a pregnancy outcome with spina bifida.

Pregnancies within a cohort of women with a diagnosis of epilepsy where the woman had been exposed to valproate as monotherapy during the first trimester were successfully identified in the GPRD. Seven cases of spina bifida were identified; one resulted in a live birth and six in a termination of pregnancy. Four of the cases were among the 225 first trimester monotherapy valproate-exposed pregnancies, two were among the 902 pregnancies unexposed to an AED and one was exposed to AEDs in polytherapy. An eight-fold increase in spina bifida risk was identified in those pregnancies exposed to valproate in monotherapy compared to those with no AED exposure (RR 8.02; CI₉₅ 1.5-43.5).

The valproate and spina bifida section of the study presented in Chapter 7 demonstrates that it is possible to identify a known teratogenic association using data recorded in the GPRD. This is of particular relevance because the

identification of the same teratogenic association formed part of the assessment of EUROCAT data to determine its suitability for use to evaluate increases in the risk of specific malformations following exposure to anticonvulsants.⁽¹¹⁾ The fact the majority of spina bifida cases were identified in pregnancies that ended in a termination highlights how essential it is that induced terminations of pregnancy are captured when serious congenital malformations, detectable at prenatal testing, are an outcome of interest. Without the inclusion of induced terminations of pregnancy it would not have been possible to replicate the association.

Identification and categorisation of pregnancy losses in the GPRD

- Having determined that it was possible to identify, by review of the free text, pregnancies that had been terminated following a prenatal diagnosis, the final objective was to determine whether this could be done in an automated manner. This involved the development of an algorithm to identify pregnancy losses in the GPRD to categorise them into those that were spontaneous, induced for a foetal medical condition and induced for other reasons.

Pregnancy losses were successfully identified in the GPRD and an algorithm was developed to categorise the type of pregnancy loss. Free text comments were requested and reviewed for a random sample of 200 pregnancy losses to evaluate the reliability of the algorithm.

For 82.0% of pregnancy losses identified in the GPRD there was sufficient information in the patient's medical record for the algorithm to categorise them as being spontaneous or induced. For induced terminations of pregnancy, in almost two-thirds of cases, it was possible to infer the reason for the termination based on the evaluation of medical codes in the patient's electronic record. Review of the additional free text comments for a sample of 200 pregnancy losses found agreement between the algorithm end point and the type of pregnancy loss for 156 cases, no additional information was available in 36 cases and the free text contradicted the algorithm for 8 out of 200 pregnancy losses.

The algorithm identified 1.8% of induced terminations as being carried out for medical reasons, 0.1% based on the level of detail in the termination code and 1.7% based on evidence of a planned pregnancy or a code for a congenital malformation.

The work presented in Chapter 8 demonstrates that it is possible to identify the full range of pregnancy losses in the GPRD. It is considered reassuring that the percentage of pregnancies identified as ending in a pregnancy loss is in line with what has been reported in the literature and this was also true for the different types of pregnancy loss. Using data recorded in the electronic records alone, however, there was insufficient information for the algorithm to determine the type of loss in around 18.0% of pregnancies and in 37.1% of induced terminations the reason was unknown. It is therefore possible that some of these could have been terminated following a prenatal malformation diagnosis but given the small percentage of all pregnancy losses that fall into this category, the total number and subsequent impact of this is likely to be small. The percentage of induced terminations carried out following a prenatal diagnosis of a congenital malformation may have been slightly overestimated using the algorithm, however the extent to which this occurs appears to be low. For these potential prenatally diagnosed cases, review of the associated free text comments is recommended to confirm malformation status. This will help avoid incorrect risk estimates that could lead to unnecessary concern or a false sense of security for women taking the product of interest.

The ability to capture pregnancy losses is a significant strength of the GPRD and is one aspect that could enable it to complement other more commonly used data sources in the area of drug safety in pregnancy. In addition to pregnancy losses the GPRD also has the potential to provide data on other adverse pregnancy outcomes; these include premature deliveries, stillbirths, pre-eclampsia and neonatal deaths. To my knowledge the recording of these pregnancy outcomes in the GPRD has not been evaluated but it is something that is going to be carried out as part of a drug safety in pregnancy study currently

being undertaken within the Pharmacoepidemiology unit at the University of Bath.

9.2 Current developments

Drug utilisation studies are important in drug safety in pregnancy research for determining priorities of study⁽⁹⁾ and for understanding the patient characteristics of user group populations to help interpret potential signals generated from sources such as registries. The presence of pregnant women regardless of exposure, in automated healthcare databases and some linked population-based surveillance systems, means there is a known denominator and this enables the prevalence of exposure to particular products to be calculated. This therefore makes these data sources particularly suitable for drug utilisation studies.

The European Commission is currently funding an Seventh Framework Programme focussing on the safety of medicine use in pregnancy.⁽¹²⁾ This study aims to evaluate the safety of products in four therapeutic drug classes (anticonvulsants, SSRIs, anti-asthma medicines and diabetes medicines) whilst also obtaining information on drug utilisation patterns. The drug utilisation aspect of the study will collate data from the GPRD in addition to databases in the Netherlands, Italy, Denmark, Norway and Wales. Some of the databases have already established themselves within the field of drug safety research^(4, 13-15) but others like the SAIL database in Wales are being utilised in this way for the first time.

This thesis has demonstrated there are many factors that need to be taken into account when using the GPRD for drug safety in pregnancy research. As the number of researchers using automated healthcare databases increases, it is important to continue sharing what is known about the intricacies of using these data sources and the importance of fully understanding the healthcare system from which the data has been collected. This knowledge can be shared via publications and commentaries in peer reviewed journals and through

workshops, symposia and meetings that are organised by the Drug safety in Pregnancy Special Interest Group that meet once a year at the International Conference of Pharmacoepidemiology.

One development that may go some way to ensuring high standards and improving the scientific rigor of studies is the creation of the European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCEPP). This aims to “improve pharmacoepidemiological research and post-authorisation safety surveillance of medicinal products in Europe by offering access to a robust network of resources working in a transparent and independent manner according to the highest scientific standards”.⁽¹⁶⁾ The ENCePP working groups have drawn up a detailed study protocol check list which it hopes will encourage researchers to consider all important aspects of study design. In addition to which, to increase transparency, all studies eligible for ENCePP approval need to have their study protocol posted to the ENCePP website before the study commences. Finally the latest developments at ENCePP include the creation of Special Interest Groups with the aim to facilitate discussions between the EMA and its Committees and ENCePP. It is possible one such special interest group might focus on drug safety in pregnancy.

9.3 Overall conclusions

The work carried out as part of this thesis, which aimed to evaluate the utility of the General Practice Research Database to act as an alternative or complement to pregnancy registries, has demonstrated that:

1. It is possible to identify all types of pregnancy outcome, including pregnancy losses, in the GPRD.
2. First trimester exposure to anticonvulsants can be determined by identifying and mapping prescription records.
3. Within the GPRD it is possible to link the mother's medical record to that of the child in over 90% of live deliveries.
4. A range of major congenital malformations can be identified from codes in the child's and mother's GPRD medical records.
5. The reliability of major congenital malformation recording is high but verification of those identified as part of future studies is recommended.
6. Using an algorithm it is possible to categorise 80% of pregnancy losses in the GPRD. Review of the free text comments is still required, however, to confirm terminations of pregnancy that followed a prenatal malformation diagnosis.
7. It is possible to carry out a risk assessment study using data from the GPRD and to identify a known teratogenic association.

The GPRD was found to have a number of key strengths for drug safety in pregnancy research when compared with some other data sources. These include the representative nature of the population captured, the availability of

the data, the ability to identify all types of pregnancy outcome, the potential for large sample sizes and the presence of a denominator population.

The capability of the GPRD is likely to be restricted, however, by small sample sizes when evaluating products used to treat less prevalent conditions, the inaccuracies associated with establishing exposure based on the issue of a prescription and a lack of information on some potential confounding factors such as over-the-counter folic acid exposure. Further validation of the algorithms used to identify pregnancies and categorise pregnancy losses are required in addition to verifying the recording of other adverse pregnancy outcomes.

At present the GPRD is most suited to use for carrying out studies of drug utilisation in pregnant women and evaluating the safety of medicines used to treat chronic and more prevalent conditions or those where there is a substantial increase in risk. It has been frequently commented that it is unlikely a single data source or study design will be sufficient for monitoring all aspects of the safety of medicine use during pregnancy.^(17, 18) This thesis has demonstrated that the General Practice Research Database has the potential to make a valuable contribution to the field of drug safety in pregnancy and could play an important role in complementing the work of other surveillance systems including pregnancy exposure registries and case-control systems.

9.4 References

1. Alberman ED. Spontaneous abortion: epidemiology. In: Stabile S, Grudzinkas G, Chard G, editors. Spontaneous abortion: diagnosis and treatment; 1992 p. 9-20.
2. Wilcox AJ, Weinberg CR, O'Connor JF, et al. Incidence of early loss of pregnancy. *N Engl J Med*. 1988;319(4):189-94.
3. Mishell D, Stenchever M, Droegemueller W, et al. Comprehensive gynecology. Missouri: Mosby-Year Book, Inc; 1997.
4. Devine S, West S, Andrews E, et al. The identification of pregnancies within the General Practice Research Database. *Pharmacoepidemiol Drug Saf*. 2010;19(1):45-50.
5. Toh S, Mitchell AA, Werler MM, et al. Sensitivity and specificity of computerized algorithms to classify gestational periods in the absence of information on date of conception. *Am J Epidemiol*. 2008;167(6):633-40.
6. Rockenbauer M, Olsen J, Czeizel AE, et al. Recall bias in a case-control surveillance system on the use of medicine during pregnancy. *Epidemiology*. 2001;12(4):461-6.
7. Wurst KE, Ephross SA, Loehr J, et al. The utility of the General Practice Research Database to examine selected congenital heart defects: a validation study. *Pharmacoepidemiol Drug Saf*. 2007;16:867-77.
8. Wurst KE, Ephross SA, Loehr J, et al. Evaluation of the General Practice Research Database congenital heart defects prevalence: Comparison to United Kingdom national systems. *Birth Defects Res A Clin Mol Teratol*. 2007;79:309-16.
9. Hernandez-Diaz S. Prescription of medications during pregnancy: accidents, compromises and uncertainties. *Pharmacoepidemiol Drug Saf*. 2006;15:613-7.
10. Laudati J, Kou T, Wilson K, et al. Comparison of methods to define lifestyle risk factors in the General Practice Research Database (GPRD). *Pharmacoepidemiol Drug Saf*. 2011;20(S1).
11. Dolk H, Jentink J, Loane M, et al. Does lamotrigine use in pregnancy increase orofacial cleft risk relative to other malformations? *Neurology*. 2008;71:714-22.
12. Morgan M, De Jong-Van Den Berg LTW, Jordan S. Drug safety in pregnancy – monitoring congenital anomalies. *J Nurs Manage*. 2011;19(3):305-10.
13. Bakker MK, Jentink J, Vroom F, et al. Drug prescription patterns before, during and after pregnancy for chronic, occasional and pregnancy-related drugs in the Netherlands. *BJOG*. 2006;113(5):559-68.
14. Gagne JJ, Maio V, Berghella V, et al. Prescription drug use during pregnancy: a population-based study in Regione Emilia-Romagna, Italy. *Eur J Clin Pharmacol*. 2008;64(11):1125-32.
15. Molgaard-Nielsen D, Hviid A. Newer-generation antiepileptic drugs and the risk of major birth defects. *JAMA*. 2011;305(19):1996-2002.
16. European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP). Available from <http://www.encepp.eu/> [Accessed 12.03.2012].
17. Mitchell A. Studies of drug-induced birth defects. In: Strom BL, editor. *Pharmacoepidemiology*: John Wiley and Sons Ltd; 2005. p. 501-14.
18. Howard TB, Tassinari MS, Feibus KB, et al. Monitoring for teratogenic signals: Pregnancy registries and surveillance methods. *Am J Med Genet C Semin Med Genet*. 2011;157(3):209-14.

Appendix I